UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FARMACIA

TESIS DOCTORAL

Estudios de RMN y dinámica molecular. Reconocimiento molecular de carbohidratos por proteínas

NMR and molecular dynamics studies on the molecular recognition of carbohydrates by proteins

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Pilar Blasco Morales

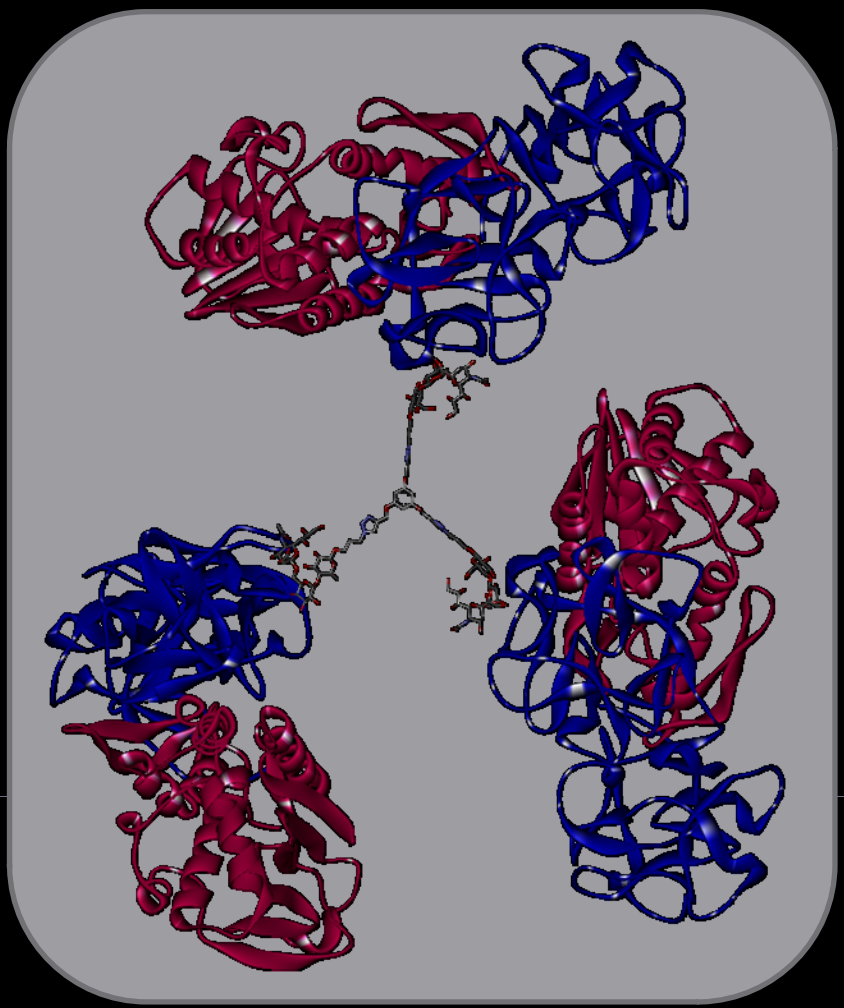
Director

Jesús Jiménez Barbero

Madrid, 2014

Estudios de RMN y Dinámica Molecular. Reconocimiento Molecular de Carbohidratos por proteínas

NMR and Molecular Dynamics Studies on the Molecular Recognition of Carbohydrates by Proteins



PhD Thesis

PILAR BLASCO MORALES

CSIC

CENTRO DE INVESTIGACIONES BIOLÓGICAS

Unversidad Complutense de Madrid
Facultad de Farmacia

Centro de Investigaciones
Biológicas(CSIC)

Programa de Doctorado

FARMACIA

ESTUDIOS DE RMN Y DINÁMICA MOLECULAR. RECONOCIMIENTO DE CARBOHIDRATOS POR PROTEÍNAS

NMR AND MOLECULAR DYNAMICS STUDIES ON THE MOLECULAR RECOGNITION OF CARBOHYDRATES BY PROTEINS

Memoria para optar al grado de doctor en Farmacia presenta:

Pilar Blasco Morales

Madrid, 2013

Director:

Dr. Jesús Jiménez Barbero

Centro de Investigaciones Biológicas

Consejo Superior de Investigaciones Científicas

A mis padres,

Agradecimientos

Rápido ha pasado el tiempo, pero la dicha ha sido buena. Quiero agradecer a mi director de tesis, el Prof. Jesús Jiménez Barbero por haberme dado la oportunidad de poder realizar mi tesis en el grupo durante estos casi tres años, por todo lo que he podido aprender gracias a este tiempo que me ha brindado y por sus consejos y ayuda constante. También quiero agradecer al Prof. Javier Cañada por todos sus buenos consejos; me ayudaron mucho a la hora del trabajo en el laboratorio.

Quiero agradecer a la Prof. Mª José Hernaiz tanto por su ayuda como tutora de mi tesis como su ayuda a nivel profesional; gracias a ella muchas cosas fueron más sencillas.

Como no, quiero agradecer a mis dos patitas con las que me encuentro fuertemente acoplada, Ángeles y Anita (Nani). Esos momentos "triplete" fueron cruciales a la hora de sacar todo este trabajo adelante. Siempre tendré un huequecito o un sofá libre para vosotras donde quiera que esté. Gracias a las dos por toda vuestra ayuda.

También quiero agradecer a mis compañeros de laboratorio; Álvaro, Javierín, Dolores, Khouzaima, Luca, Ana, Carmen y Silvia. La vida en el laboratorio no es solo 100% trabajo, también es encontrarte a gusto con la gente que tienes alrededor. ¡Por todo eso gracias chicos!

También quiero dar las gracias a toda la gente que han estado durante mi estancia en el CIB; Álvaro, Luigi, Virginia, Filipa, Faina, Roberta, Matteo, Carlottiña, Ángel, Alessandra y Javier. También a mi querida Blanquita que me enseñó todo lo que sabe (que no es poco) los primeros meses cuando llegué al laboratorio.

Gracias a mis compis de la Uni, en especial a Ana, Salo, Ade, Blanca, Dafne, Judith y Patri. Esos momentos juntas de vez en cuando me animan muchísimo.

También a mis amigos de toda la vida, tanto los que están cerca, como los que están lejos; entre ellos Nuria, Noelia, Prima Rebe, Isma, Alberto, Aye, Heather, Ángela, Roci, Mamen y a mi cuñadísima Mari. En lo malo, pero sobre todo en lo bueno, siempre estuvisteis y quiero que estéis. Gracias.

También a toda mi familia, a mi hermano y sobre todo a mis padres, son mi fuente de motivación e inspiración, y sin ellos, no habría llegado a ser ni la milésima parte de lo que soy ahora. Gracias por enseñarme a valorarme y por formarme como la persona que soy a día de hoy. Todo lo que soy es gracias a vosotros.

Por supuesto, gracias a Sergio, mi máximo apoyo continuo tanto en lo bueno como en lo malo, por entender y comprender todo lo que supone el trabajar en el mundo de la ciencia, por su cariño y su paciencia; en dos palabras, mi complemento.

Quiero también dar las gracias a mis colaboradores; los grupos de la Dra. Dolores Solís del IQFR, del Prof. Lokesh Joshi en el NUI en Irlanda, de la Prof. Laurence Mulard en el Instituto Pasteur en Francia.

La realización de esta tesis ha sido posible gracias a la financiación del EU's Seventh Framework Programme bajo el proyecto: ***GlycoHIT-*** *Glycomics by High throughput Integrated Technologies (HEALTH.2010.1.1-3, Ref. 260600).*

La realización de esta tesis también fue posible gracias a la financiación de Ministerio de Economía y Competitividad, bajo el proyecto: RMN Y RECONOCIMIENTO MOLECULAR. INTERACCIONES PROTEINA CARBOHIDRATO Y ÁCIDO NUCLEICO-CARBOHIDRATO (CTQ2012-32025).

Resumen

El reconocimiento de carbohidratos por lectinas (proteínas que selectivamente reconocen carbohidratos sin actividad enzimática o inmunológica) es un suceso ubicuo en los organismos vivos. De hecho, median variedad de procesos biológicos clave. Por esta razón, el interés por estudiar carbohidratos desde esta perspectiva ha aumentado en los últimos años. Desde un punto de vista estructural, una gran cantidad de técnicas biofísicas han sido empleadas para obtener información detallada acerca de éstos procesos de reconocimiento molecular esenciales. En particular, hemos destacado el uso combinado de experimentos de RMN con simulaciones de Dinámica Molecular y protocolos de modelado molecular como herramientas clave para diseccionar, a escala atómica, los elementos de unión clave de ligandos relevantes desde un punto de vista biomédico en sus superficies del receptor específico.

Por consiguiente, en esta tesis, hemos empleado una combinación de RMN-STD, trNOESY y simulaciones de dinámica molecular como poderosas herramientas para obtener información importante acerca del reconocimiento molecular de oligosacáridos sintéticos frente a dos anticuerpos específicos IgG C7-37 e IgM G19-2. Nuestro estudio ha dado evidencias experimentales de la importancia individual de los diferentes residuos, dependiendo de su posición en la secuencia oligosacarídica, también para oligosacáridos mas extensos, y su relevancia en el proceso de reconocimiento molecular. La información obtenida a partir de los experimentos de RMN-STD evidenció la directa implicación del residuo de RhaC seguido por sus residuos vecinos (Rha*p*B-**Rha*p*C**-GlcNAc*p*D) en todos los oligosacáridos sintéticos en presencia de ambos anticuerpos. Experimentos de RMN-trNOESY también dieron información estructural de los valores de Φ/Ψ mayoritarios para cada enlace glicosídico, además de el equilibrio conformacional existente en el estado unido. Este equilibrio ha sido comparado con el existente en el estado libre. Los datos obtenidos han indicado que estos procesos de reconocimiento molecular conllevan una selección conformacional. El análisis de las superficies moleculares obtenidas para los confórmeros acomplejados mostró que las mayoritarias áreas no-polares del ligando están cerca de la superficie del anticuerpo.

Cuando los N-glicanos interaccionan con lectinas, la naturaleza del epítopo objetivo estrictamente depende de la naturaleza de la lectina estudiada, pudiendo seleccionar distintos determinantes de la cadena sacarídica del *N*-glicano con exquisita especificidad, dependiendo de la presentación relativa de los diferentes residuos y de la arquitectura del sitio de unión. Hemos enfatizado la importancia de la coexistencia de sitios de unión de diferentes dominios de lectina para conseguir unión al glicano comparando un dominio único de heveina y WGA, un tetrámero. La unidad monomérica no reconoce el extenso glicano, con la ramificación en el residuo de βMan siendo el elemento clave de descarte de la interacción. Por el contrario, la lectina multidominio WGA se une al extremo reductor de ambos N-glicanos sialilado y no sialilado admitiendo la posibilidad de obtener interacciones estabilizantes entre-sitios de uno de los dominios de heveína vecinos. Mediante estudios con VAA se obtuvo información acerca de cómo la α-sialilación 2,6 puede ser acomodada por esta lectina, generando un incremento de la unión en *N*-glicanos naturales cuando lo comparamos con el *N*-glicano no sialilado correspondiente. La unión al extremo terminal LacNAc del *N*-glicano sialilado es detectable, obteniéndose la referencia control para la unión positiva.

También en esta tesis, hemos demostrado la capacidad de cuatro oligosacáridos relacionados con antígeno glycotope (relacionado con procesos de respuesta inmune) de interaccionar con diferentes fragmentos de anticuerpo scFv que se unen al glycotope. Usando experimentos de STD, hemos demostrado que la presencia del residuo 3-O-α-*DGal* A' es decisiva en el proceso de unión.

Finalmente, también hemos demostrado que una variedad de glicodendrímeros funcionalizados con lactosa y con 6'sialil-lactosa pueden ser fácilmente acomodados por Viscumina (una lectina modelo tóxica) mejorando la unión. Experimentos de RMN asistidos por modelado molecular permitieron la identificación de los rasgos de unión clave entre el azúcar y la lectina. El protocolo de modelado ha permitido sugerir un modelo estructural de las posibles interacciones multivalentes de estos ligandos con esta lectina.

Abstract

The recognition of carbohydrates by lectins (proteins that selectively recognize carbohydrates without enzymatic or immunological activity) is a ubiquitous event in living organisms. Indeed, they mediate a variety of key biological processes. For this reason, the interest to study carbohydrates from this perspective has increased in the last years. From the structural viewpoint, a number of biophysical techniques have been employed to obtain detailed information about these essential molecular recognition processes. In particular, we have stressed the combined use of NMR experiments with MD simulations and molecular modeling protocols as key tool to dissect, at atomic resolution, the key binding elements of biomedical relevant ligands at their specific receptor surfaces.

Thus, in this thesis, we have employed a combination of STD, trNOESY-NMR and MD simulations as a powerful tool to provide important information about the molecular recognition of synthetic oligosaccharides towards two specific IgG C7-37 and IgM G19-2 antibodies. Our study has provided experimental evidences on the individual importance of the different residues, depending on its position at the oligosaccharide sequence, also for longer oligosaccharides, and their relevance for the molecular recognition process. The information obtained from STD-NMR experiments have evidenced the direct involvement of residue RhaC followed by its neighbor residues (Rha*p*B-**Rha*p*C**-GlcNAc*p*D) in every synthetic oligosaccharide in the presence of both mAbs. trNOESY-NMR experiments have also provided structural information on the major Φ/Ψ values for each glycosidic linkage, as well as the existing conformational equilibrium in the bound state. This equilibrium has been compared to that existing in the free state. The obtained data have indicated that these molecular recognition processes involve conformational selection processes. The analysis of the obtained molecular surfaces for the bound conformers showed that the major non-polar areas of the ligand are close to the antibody surface.

When large *N*-glycans interact with lectins, the nature of the target epitope strictly depends on the nature of the tested lectin, which can select distinct determinants of the complex-type *N*-glycan saccharide chain with exquisite specificity, depending on the relative presentation of the different residues and the architecture of the binding site.

We have underscored the importance of the coexistence of binding sites from different lectin domains for achieving glycan binding by comparing a single hevein domain and WGA, a tetramer. The monomeric unit does not recognize the large glycan, with the branching at the βMan residue being the key structural element for precluding the interaction. In contrast, the multidomain lectin WGA binds the terminal reducing end of either non- or sialylated-*N*-glycan owing to the possibility of achieving additional inter-site stabilizing interactions from one of the neighboring hevein domains. Work with VAA has yielded insights into how α2,6-sialylation can be accommodated by this lectin, providing enhanced binding in natural *N*-glycans when compared to the non-sialylated LacNAc-terminated analogue. The results for the MAL–MAH mixture have permitted definition of the role of the α2,6-linked sialic acid residue in blocking binding of the corresponding sialylated-*N*-glycan. The binding to the LacNAc terminus of the non-sialylated-*N*-glycan is detectable, providing the control reference for positive binding.

Also in this thesis, we have demonstrated the capability of four oligosaccharides related to the glycotope antigen (related to immunodeficiency processes) to interact with different scFv glycotope-binding antibody fragments. Using STD-NMR experiments, we have shown that the presence of 3-O-α-*DGal* A' moiety is decisive for the binding process.

Finally, we have also shown that a variety of lactose- and 6'sialyllactose-functionalized glycodendrimers can be easily accommodated by Viscumin, a model toxic lectin, providing enhanced binding. NMR experiments assisted by molecular modeling permitted the identification of the key sugar-lectin binding features. The modeling protocol has permitted to suggest a structural model for the possible multivalent interactions of these ligands with this lectin.

Chapter 1. Introduction

1.1 Molecular recognition. Carbohydrates and Lectins

1.1.1 Molecular Recognition Processes

Molecular recognition is at the heart of life events. The interaction between molecules is of paramount importance for a variety of processes related with life and death. Carbohydrate molecules, which mainly exist conjugated in glycoproteins or glycolipids, are the key point of molecular recognition processes since they are at the periphery of the cell. Indeed, they mediate a variety of intra-, extra-cellular and pathological events, as well as cell-cell and cell-matrix interactions. Modifications of cell glycosilation also occur during inflammation and cancer[1]. Therefore, the Glycobiology field, somehow between chemistry and biology, with implications in biomedicine, is a challenging and active research field. In this context, the study at the maximum possible resolution of the role of carbohydrates into key biological interactions is of paramount important to completely understand vital processes and to open the possibility of their modulation.

1.1.2 Carbohydrates as recognition molecules

Nowadays, it is well described that carbohydrates play an important role in the specific molecular recognitions of different entities, of different size and complexity, such as lectins, enzymes, virus, and antibodies[2]. As mentioned above, they mediate diverse cellular activities, such as cell recognition, growth, apoptosis and immune response[3].

Carbohydrates constitute one of the most abundant and complex types of biomolecules in nature, and constitute markers for molecular recognition processes. They display a huge potential to encode biological information, given their complex spatial structure[4] They are composed by simple moieties (monosaccharides), which may be connected each other by a great variety of linkage types. They can form linear sequences or sophisticated branched structures[5]. Moreover, the introduction of substituents, as phosphate, sulfate groups[6] or/and acetyl[7] groups at the hydroxyl groups highly increases the carbohydrate complexity. This structural diversity allows carbohydrates to specific recognize receptors with exquisite efficiency. Therefore, the

knowledge of the atomic three dimensional (3D) perspective of the molecular recognition process between saccharides and their receptors is a topic of major interest for designing carbohydrate-based therapeutic agents.

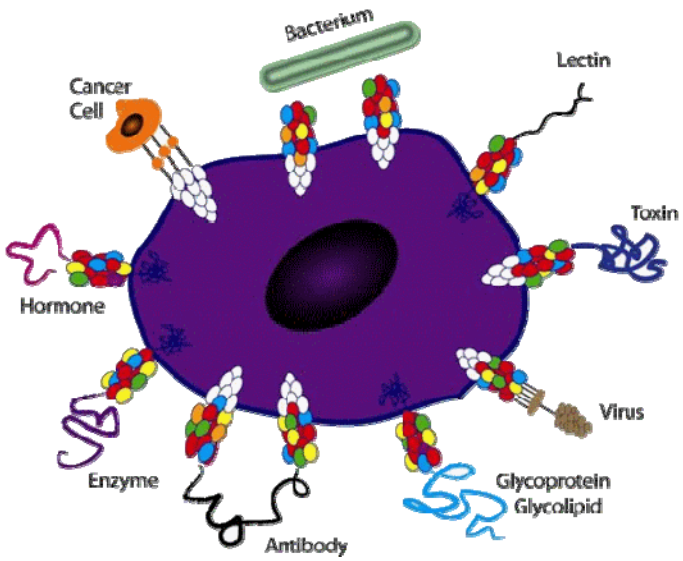


Figure 1.1 Cell surface carbohydrates are recognition molecules used in specific binding and interactions among cells, pathogens, and molecules.

1.1.3 Protein-carbohydrate interactions: CH-π, polar interactions, hydrogen bonding, and the role of water

In the last years, different investigations have deeply looked at the fine structural details of protein-carbohydrate interactions. Advances in different biophysical and spectroscopic techniques, as X-Ray crystallography and NMR spectroscopy, complemented by quantitative data obtained through titration microcalorimetry (TMC)[8] have allowed the access to detailed information on the atomic structure of protein-carbohydrate complexes[9] and permitted to speculate on the forces involved in the interaction. From the structural perspective, hydrogen bonds (H-bonds) and van der Waals forces (often including packing of a hydrophobic sugar face against the aromatic amino acid side chains) are the major factors (Figure 1.2) involved in the binding.

This fact is not unexpected, since from a general molecular recognition viewpoint, the typical contributions to the Gibbs' free energy of ligand binding

(Equation 0) are originated from hydrogen bonding, van der Waals forces, entropy contributions and from the solvent reorganization. In particular, the role of the alterations of the geometry and motion dynamics of the receptor, ligand and/or solvent molecules should be highlighted[10] (Fig. 1.2).

Eq. 0

$$\Delta G^{O} = \Delta H^{O} - T\,\Delta S^{O} = RT\,ln\,K_a$$

This equation also encodes a fundamental relationship between the enthalpy and entropy factors that may be attributed to the participation of weak intermolecular forces. It has been well demonstrated that in many cases, when comparing different ligands to a given receptor, an increase in the binding enthalpy (ΔH^{O}) is will inherently balanced by an entropic (ΔS^{O}) penalty (or vice versa). This phenomenon is noted as the enthalpy-entropy compensation effect[11].

There are three prevailing models for describing the molecular basis of enthalpy/entropy compensation in the binding of oligosaccharides to proteins[12]. The Lemieux's group first reported compensation in the binding of oligosaccharides to proteins[13]. Initially, the observed compensation was attributed to the loss of protein degrees of freedom. Later, Lemieux[14] suggested the existence of differential solvation effects to be responsible for the compensation, a hypothesis closer to that of Lumry and Rajender[11]. Carver suggested a third model, where the loss of conformational entropy by the ligand on binding provided the major source of the entropic barrier, at least for oligosaccharide binding to lectins[15]. All three mechanisms are undoubtedly operative to some extent. The question remains, however, as to which effect is predominant. A complete answer could possible derive by only an accurate analysis of the behavior of each participant at the molecular recognition process.

Because of the amphiphilic character of oligosaccharides, the recognition process involves forces of different nature. These types of interactions can be clearly seen in the available X-Ray and NMR structures of carbohydrate-protein complexes (Figure 2). Polar groups of the sugars are bound by H-bond donors and acceptors in the protein backbone and polar side chains. Apolar groups (especially formed by CH groups of the pyranose) are complemented by nonpolar surfaces of the proteins. In fact, the CH vectors tend to pack against the aromatic residues at the receptor side chains[16] (Figure 1.2). NMR studies have shown the presence of interactions between certain hydrogens

belonging to the non-polar faces of the sugar ring and aromatic residues of protein side chains. Indeed, these studies have linked the strength of the interaction to the size and the electron richness of the arene[17].

Depending on the conformation of the carbohydrates, an arrangement of three C-H groups with axial orientation on one face of the sugar can effectively interact with the surface of a properly positioned aromatic residue. The strength of this interaction is relatively weak, since a unique CH-π interaction may worth less than 1 kcal/mol[18]. However, the simultaneous interaction between several CH groups with one aromatic system enhances the strength of the interaction[19]. Clearly, the orientation of the aromatic side-chain is determinant for the specificity of the binding.

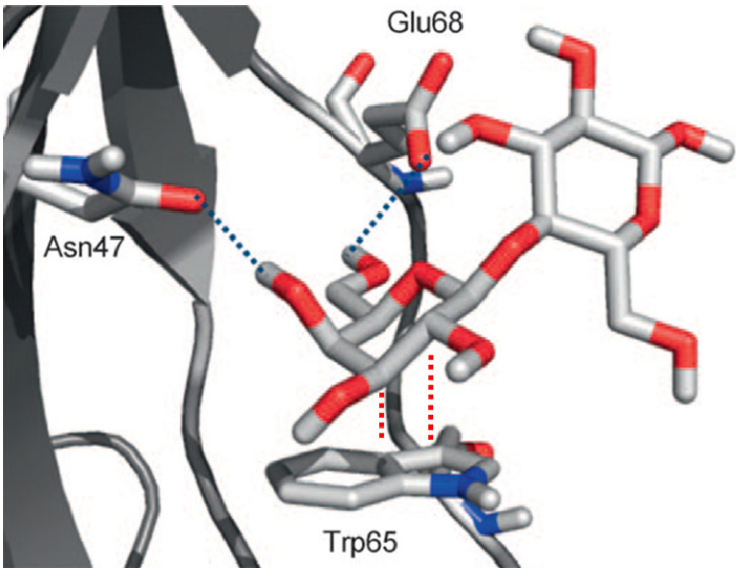


Figure 1.2 General description of the interactions involved in protein-carbohydrate binding. Schematic representation of hydrogen-bond and CH-π interactions between aminoacid residues of galectin-1 and lactose[10]. Besides hydrogen bonds between several hydroxyl groups of the saccharide and protein polar side chains, stacking interactions between several sugar CH bonds and aromatic rings endow the complex with stability.

Water molecules also play an important role in the molecular recognition process. There has been a large debate on this topic and, indeed, Lemieux and co-workers[14b] suggested that the reorganization of water molecules upon binding provides the driving force for the carbohydrate complex formation. In the receptor, the binding pocket is occupied by water molecules that stabilize the orientation of the protein residues, later involved in the recognition process, through a network of interactions.

When the positions of these waters and aminoacids before and after complex formation with the sugar are compared, some of these water molecules in the free state seem to occupy the positions of the hydroxyl groups of the sugar in the complex[20]. Thus, in this way, no displacements of aminoacid side chains are necessary. Also, those water molecules that are directly involved in the interactions between the protein and the sugar display restricted motions, compared to other water entities in bulk solution and, therefore they may not be able to adopt the suitable orientation to form hydrogen bonding with other vicinal waters. When the complex is going to be formed, the water molecules that solvate the polar surface of the protein binding pocket have to be displaced to leave the space for the sugar to enter. In the same way, those water molecules solvating the polar surface of the carbohydrate are released into the bulk solution. All these processes display key enthalpy and entropy requirements. It is evident that the reorganization of the solvent implicates many molecules, a multivalency process that leads to a strong contribution to the thermodynamics of the system. In fact, according to Lemieux[14b], the release of the geometry restrains on the receptor-bound water molecules may provide a key enthalpy driving force to the binding process ("hydrophobic effect").

1.1.4 The receptors: Lectins and Antibodies

Lectins are proteins that bind carbohydrates with a high selectivity, reversibility and efficacy. Lectins are found in many cellular layers of a large number of organisms, and their localization reflects their diversity of function. They used to appear in the cell surface regulating cellular adhesion and migration processes and, in some cases, helping to the cellular metabolism control[21]. Investigation of lectins and their role in cell recognition, as well as the application of these proteins for the study of carbohydrates in solution and on cell surfaces, are making marked contributions to the advancement of Glycobiology[1b]. They are not neither antibodies nor enzymes. They are proteins with non-immunogenic origin and do not modify the carbohydrate that binds, but can be as specific as the interaction between antigens and antibodies or substrates and enzymes[21b, 22]. Lectin-monosaccharide interactions, however, are relatively weak with dissociation constants often on the order of micromolar to millimolar range[21a, 23].

Lectins were firstly identified in plants (See Table 1.1). Only later were found in microorganisms and animals. It can be safely assumed that only a small fraction of the total number of lectins that exist in nature have been now identified. They can be identified based on functional assays (for instance, hemagglutination) or by amino acid sequence homology (putative lectins) with known lectin sequences. Nowadays, only a small fraction of the discovered lectins have been carefully characterized with respect to protein structure, binding affinities for extended carbohydrates, and binding thermodynamics. This fact is changing, and there is a great interest in the elucidation of the function of endogenous lectins, since it might be correlated with a deeper knowledge of many diseases and their associated pathogenesis.

1.1.4.1 *Plant lectins*

The most known members of the diverse group of plant lectins are shown in Table 1. The leading position is held by concanavalin A, the classical Man/Glc-binding plant lectin from *Jack beans*. Because of numerous chemical modifications, concanavalin A is very useful for chemical, biochemical and biomedical applications[4]. Moreover, this lectin was the first one to be characterized by X-Ray diffraction methods.

In general, they show several functionalities, from plant defence from pathogenic attacks to cell wall extension, but also they are involved in mitogenic simulation, transport of carbohydrates, mobilization of storage material, and as starting points of the symbiosis processes in leguminous plants[1c].

- HEVEIN

There are a group of lectins with vegetal origin, with small size, that bind to N-Acetyl-glucosamine (GlcNAc) oligomers. Indeed, chitin is a polysaccharide [24] that is the principal component of insects and crustacean's exoskeleton and also is present in the cellular wall of a large number of microorganisms[25]. These lectins are characterized by a common structural motif, which is referred as the hevein domain[26]. In fact, because of their intrinsic structural properties, joined to their relative easy availability, hevein

has been employed as a model to study the mechanism, origin and structural properties involved in protein-carbohydrate molecular recognition processes.

Plant species and name/ abbreviation of lectin	Family	Mono- or disaccharide specificity	Comments
Canavalia ensiformis (concanavalin A, ConA)	Leguminosae	Man/Glc	cheapest and most popular lectin; first lectin isolated by crystallization and demonstrated to interact with carbohydrate (see text and Table 2 for details)
Ricinus communis (ricin)	Euphorbiaceae	Gal	ribosome-inactivating protein, type II (RIP II), used for generating immunotoxins; biohazard
Triticum vulgare (WGA)	Gramineae	$(GlcNAc)_{1-3}$, Neu5Ac	potential function in plant defence mechanisms
Phaseolus vulgaris (PHA)	Leguminosae	no simple carbohydrate known	isolectin L_4 is a strong mitogen for T-lymphocytes, isolectin E_4 is a strong erythrocyte agglutinin (see Figure 1B for chromatographic isolectin separation); distinguish between bisected and nonbisected *N*-glycans; cause of severe gastrointestinal irritation when ingested in insufficiently cooked beans
Glycine max (SBA)	Leguminosae	GalNAc/Gal	cell sorting, bone marrow purging
Pisum sativum (PSA)	Leguminosae	Man/Glc	binding of *N*-glycans enhanced by core fucosylation
Viscum album (VAA, viscumin)	Viscaceae	Gal	RIP II used for generating immunotoxins, constituent of proprietary mistletoe extracts (immunomodulatory and growth stimulatory for tumor cells in vitro and in vivo at low doses; see text for details)
Arachis hypogaea (PNA)	Leguminosae	Gal, Galβ3GalNAcα (TF-antigen)	very popular in histochemistry; separates immature from mature thymocytes
Lens culinaris (LCA)	Leguminosae	Man/Glc	binding of *N*-glycans enhanced by core fucosylation; lymphocyte mitogen
Dolichos biflorus (DBA)	Leguminosae	GalNAcα3GalNAc, GalNAc	cell sorting, agglutinates blood group A erythrocytes
Griffonia simplicifolia (GSA-I)	Leguminosae	Gal/GalNAc	isolectin GSA-I-A_4 agglutinates blood group A erythrocytes, isolectin GSA-I-B_4 blood group B erythrocytes
Griffonia simplicifolia (GSA-II)	Leguminosae	$(GlcNAc)_n$	insecticidal activity, potential defence role
Artocarpus integrifolia (jacalin)	Moraceae	Gal (Man, TF-antigen)	used for isolation of IgA_1 and mucins, mitogenic for $CD4^+$ T-cells
Solanum tuberosum (STA)	Solanaceae	$(GlcNAc)_n$	potential function in plant defence mechanisms
Galanthus nivalis (GNA)	Amaryllidaceae	Man	does not bind Glc as the Leguminosae lectins do, application for insect and nematode defence in transgenic crop plants tested, antiretroviral activity in vitro, selective agglutination of rabbit but not human erythrocytes
Ulex europaeus (isolectin UEA)-I	Leguminosae	L-Fuc	agglutinates blood group 0(H) erythrocytes; selective marker for endothelial cells of primates
Erythrina corallodendron (ECA)	Leguminosae	Galβ4GlcNAc, Gal, GalNAc	mitogen for human lymphocytes
Vicia faba (VFA)	Leguminosae	Man/Glc	binding of *N*-glycans enhanced by core fucosylation
Sambucus nigra (SNA)	Caprifoliaceae	Neu5Acα6Gal/GalNAc, (Gal/GalNAc)	probe for sialylated glycoconjugates, e.g. in thymocyte differentiation
Abrus precatorius	Leguminosae	Gal	RIP II used for generating immunotoxins
Lotus tetragonolobus (LTA)	Leguminosae	L-Fuc	agglutinates red cells of blood group 0(H), instrumental to the definition of α-L-fucose as a crucial 0(H) epitope (see Table 2)
Lycopersicon esculentum	Solanaceae	$(GlcNAc)_n$	potential function in plant defence mechanisms; marker of endothelium of small vessels in rats
Phaseolus lunatus limensis	Leguminosae	GalNAcα3[Fucα2]Gal, GalNAc	agglutinates blood group A erythrocytes
Datura stramonium (DSA)	Solanaceae	$(GlcNAc)_n$	potential function in plant defence mechanisms
Maackia amurensis (MAA)	Leguminosae	Neu5Acα3Gal/GalNAc	probe for sialylated glycoconjugates
Phytolacca americana (PWM)	Phytolaccaceae	GlcNAc	known as pokeweed mitogen; detected in 1969 in the course of investigating a fatality associated with ingestion of pokeweed berries
Bauhinia purpurea (BPA)	Leguminosae	GalNAcβ3GalNAc, GalNAc	enrichment of B lymphocytes, isolation of T cells producing Il-2
Urtica dioica (UDA)	Urticaceae	$(GlcNAc)_n$	antifungal activity
Hevea brasiliensis (hevein)	Euphorbiaceae	$(GlcNAc)_n$	antifungal activity; allergen in rubber products of poor quality
Maclura pomifera (MPA)	Moraceae	T-antigen > Tn-antigen	mitogen for lymphocytes

Table 1.1 Plant Lectins: origin, families, specificities, and some general features.

The hevein domain contains around 40-50 aminoacids and is characterized by the presence, in conserved positions, of 6 or 8 cystein moieties that form three or four disulfide bridges. This name came from hevein itself, a 43 aminoacids polypeptide, which is extracted from the latex gum of *Hevea brasilensis* plant[27].

Regardless of its defense properties in the origin plant, hevein as well as other homologue proteins present in other plants are directly related with latex allergy[28]. Hevein domains can be founded isolated, as for example in hevein itself or in pseudohevein, or being part of bigger proteins, and even covalently grouped in two or more domains. For example, the nettle lectin (UDA, *Urtica dioica agglutinin*) and wheat germ agglutinin (WGA), show two and four hevein domains, respectively.

- WGA

Wheat-germ agglutinin is a lectin capable of agglutinating erythrocytes and other types of cells[29]. As mentioned above, WGA is a dimeric lectin, where each subunit consists of an assembly of four hevein domains. Therefore, WGA is composed of eight hevein domains[30]. It has been described to be specific for terminal Neu5Ac and GlcNAc moieties[31], and also longer GlcNAc-containing oligosaccharides as N,N'-diacetylchitobiose ($(GlcNAc)_2$)[29a, 32] by using a variety of techniques, including X-ray crystallography[30] and NMR spectroscopy[33]. The deduced structures reveals that the primary binding site is constituted by three conserved aromatic residues from one hevein domain together with further polar residues (See Fig. 1.3; S114 and E115) from the neighbouring hevein domain. All together contribute to the stabilization of the complex formed, providing structural evidence for an intra-protein cooperation of lectin domains.

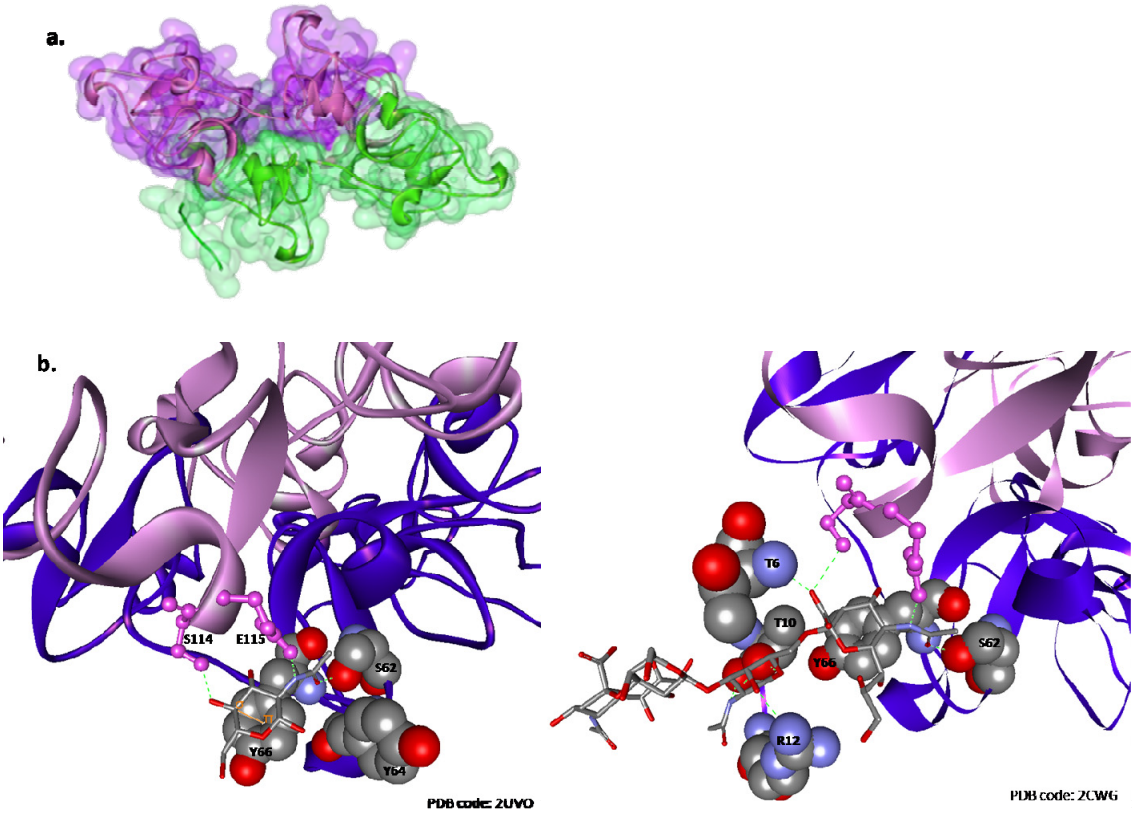


Figure 1.3 (a) Quaternary structure of WGA (*Wheat-Germ Agglutinin*). (b) Representation of GlcNAc (left) and 6'Sialyllactose (right) recognition by WGA with the key aminoacid residues involved in the binding highlighted.

- VAA

The major lectin from mistletoe extract[34], viscumin (VAA, *Viscum album agglutinin*, also referred as ML-1), is a heterodimeric lectin composed of a toxic "A" chain and a galactose-binding "B" chain. There are many toxic lectins with this construction in the plant kingdom, the best studied being ricin (RCA-II), from castor beans. A single RCA-II molecule is capable of inactivating at least 50 ribosomes in ten minutes. However, more than a citotoxic agent, viscumin is known as an immune system activator, with considerable biopharmaceutical interest. Although the structure, enzymatic activity, and carbohydrate binding ability is similar to those of type II ribosome inactivating proteins (RIPs)[34-35], nonetheless viscumin presents some peculiar differences that makes it unique.

Although viscumin is formed by the two A and B chains mentioned above, recent electrophoresis studies have revealed that these are not covalently linked by a

disulfide bond in all molecules[35a]. Viscumin forms a dimer in solution over a broad range of protein concentrations (0.01-25 mg/ml)[36]. These VAA solutions exhibit non-ideality, self-association, and poly-dispersity, as evidenced by sedimentation equilibrium and velocity experiments. When diluting the concentration down to 0.01 mg/ml, the protein behaves like a single ideal component, with the molecular mass of the dimer. At very low concentrations, viscumin monomers are also detected (concentrations <3μg/ml). However, the possible formation of non-covalent tetramers $[AB]_2$ (Figure 1.4) and the existence of a concentration-dependent dimer-tetramer equilibrium has also been reported[35a], being its coexistence also assessed by electron microscopy[37].

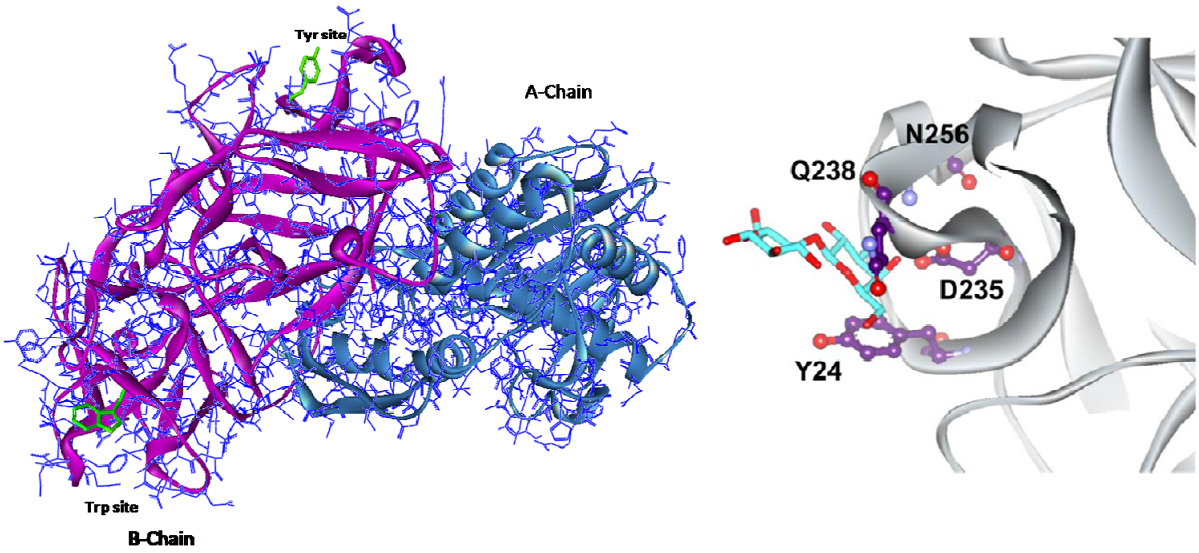


Figure 1.4 (left) Quaternary structure of the AB dimer of viscumin (PDB 1CE7). The side chain of Trp and Tyr key aminoacids are represented in sticks, as a way to indicate the two carbohydrate binding domains. (right) Representation of the lactose recognition by Viscumin.

Viscumin's chain A chain is formed by 254 amino acids, with a molecular weight of approximately 29 kDa and a pI of 6.11 (Swiss-Prot P81830). It contains a solely Cys (247) residue, which is implicated on the inter-chain disulfide bond with chain B. On the other hand, the B chain is formed by 264 amino acids, and also has an approximate 29 kDa molecular mass. In the NMR experiments that will be shown below, VAA is present in solution as a tetramer (concentration >20μM), formed by non-covalent interactions between the domains of the adjacent B chains[38], as in the ABBA tetramer.

This geometry causes the Trp CRD in both chain's (1α subdomains) to be located only 15 Å apart from one another, at the tetramer interface, with low accessibility. On the other hand, the Tyr binding site at the 2γ subdomain are present at the other protein side, separated by 87 Å. ITC and laser photo chemically induced nuclear dynamic polarization (CIDNP)[39] studies, covering a broad range of lactose/VAA ratio have evidenced that, when the lectin is in solution as a dimer (from 0.01-25 mg/ml concentrations), the Trp site in VAA is only partially occupied. The close distance between the Trp sites inevitably imposes steric restrictions on sugar binding at these loci.

Viscumin recognizes galactose specifically[40] and, oppositely to ricin, shows very low affinity towards GalNAc[40b]. The binding to galactose-ending oligosaccharides and mimetics takes place through key interactions with the galactose residue[41]. In lactose, deoxygenation at C-2 position of the Gal unit results in a considerable affinity loss (approximately 8 times), which suggest a local interaction between this hydroxyl group and the lectin. The affinity is not regained when analyzing the 2'-O-methyl derivative, which could indicate that this hydroxyl acts as an H-bond donor to the protein. Alternatively, this lack of affinity could also be due to the existence of steric factors precluding the proper docking of a methyl group[42]. The essential contribution to binding arises from the C-3' and C-4' positions. In both cases, any substitution causes a great decay on the affinity strength[42]. Nevertheless, only a slight affinity diminishing (approximately two times) is observed after dehydrogenation at C-6'. Besides, the affinity is regained after the introduction of a fluorine atom or an O-methyl group at that position, suggesting that OH-6' has H-bond acceptor properties[40a, 42] and that there is room for further extension at this region. Although the binding force of the association of viscumin with a Gal-containing ligand is generated by the non-reducing end, the residue attached to the Gal moiety as well as the anomeric configuration of that glycosidic linkage displays a decisive influence on the observed affinity [40b, 43].

Previous studies showed no significant affinity variations when large substituents were introduced at position Glc C-3 of lactose. Hydrogenation at C-6 also revealed a minor diminishing of the affinity, suggesting a weak contact of the OH-6 with the protein. Furthermore, in addition to β-linked disaccharides, such as Galβ(1-2)Galβ(1-R) and Galβ(1-3)Galβ(1-R), the lectin appears to bind well to terminal non-reducing Galα(1-3)Gal sequences of complex-type oligosaccharides and to the P1-

characteristic Galα(1-4)Gal[40, 44]. The binding domain has also been postulated to be fairly extended, capable of accommodating additional residues. In fact, VAA presents high affinity for those oligosaccharides in which the terminal Gal moiety is further sialylated at O-6, giving rise to a α(2-6)-linked saccharide. In fact, it has been suggested that VAA's "true receptors" are complex sialoglycoproteins and gangliosides holding the terminal Neu5Acα(2-6)Galβ(1-4)GlcNAc sequence[45].

The biological responses in animals and humans to viscumin action include, among others, the modulation of genetic expression, induction of cytokine secretions, release of immune-competent cells from the bone marrow, modulations of leukocytes phagocytic activity[46], protein synthesis inactivation and cellular apoptosis. Thus, viscumin acts both as a cytotoxic agent and a biologic response modulator[47].

In all cases, these processes are set in motion as a consequence of the toxin binding to specific cellular receptors. The access inside the cell may occur by several endocytic means that internalize the viscumin-receptor complex. In the initial recognition and internalization process by 3T3 cells, viscumin interacts with receptors located at the cell periphery, being then enclosed in small endocyclic vesicles present all over the cell surface, with tendency to accommodate them in cellular interior[48].

RIPs have relevant biomedical interest due to their potential use for treatment of several human diseases. To date, its most noteworthy use lies on the production of immunotoxins ant other conjugates for AIDS and cancer experimental therapy[49]. Regrettably, their use still presents clinical problems related to its toxicity. Thus, there is a significant need of insightful studies focused on the structure and ligand recognition mechanism of these proteins. This knowledge could allow the design of new alternatives, which could display a high selectivity for the therapeutic goal, raising the minimum possible side effects.

- MAL

Maackia amurensis hemagglutitin, the lectin from the seeds of *Maackia amurensis* leguminous plant, is a 26 kDa protein that is known as a strong mitogenic activator against human peripheral lymphocytes[50]. The first interaction studies showed that MAL-bound glycans with α2-3-Gal-linked sialic acid residues were efficiently recognized, as in the Sialα(2-3)Galβ(1-4)GlcNAc of N-linked glycans[51]. However, it

has been recently found that MAL can also bind to the corresponding non-sialylated glycan[52]. In our interaction studies, we have used the mixture of the two lectin species present in the seed of *M. amurensis*, MAL and MAH.

1.1.4.2 *Carbohydrate-recognizing antibodies*

Antibodies are γ-globulin proteins that can be found in blood and serum of vertebrates. These molecules are used by the immune system to recognize and neutralize foreign systems, such as virus and bacteria. In fact, as mentioned above, bacteria and viruses contain unique carbohydrates on their surfaces, which are then recognized by the key antibodies thus activating the human immune system[53].

The basic structure of an antibody consists of two identical large heavy chains (H chains) accompanied by two identical small light chains (L chains), which are covalently attached by a disulfide bond to each other (see Fig 5). They can be grouped into different isotypes based on the type of heavy chain. For mammals, there are five different antibody isotypes (IgG, IgA, IgM, IgD, and IgE). They display different functionality, thus allowing the recognition of different foreign objects (Fig 1.5).

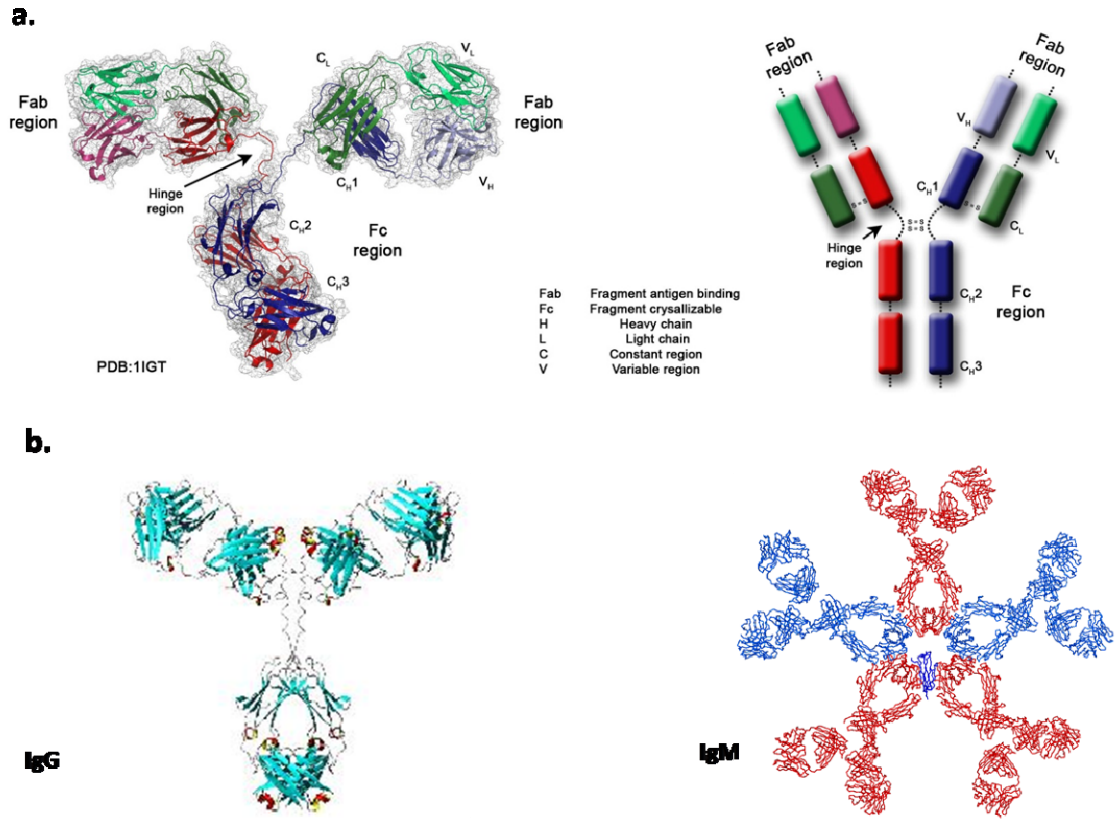


Figure 1.5 (a) Schematic structure representations of a typical IgG antibody with the different parts detailed. (b) Structure representation of IgG and IgM antibodies.

All antibody molecules share the same basic structural characteristics but display remarkable variability in the regions that bind their antigens. The part of the antigen that is highly specifically recognized by the specific antibody is called the epitope. The variability of the antigen-binding regions (or variable regions, VH and VL) accounts for the capacity of the different antibodies to bind a tremendous number of structurally diverse antigens. There are more than 10^7, perhaps as many as 10^9 different antibody molecules in every individual, each with unique amino acid sequences in their antigen-combining sites. The effector functions and common physicochemical properties of the antibodies are associated with the non-antigen-binding portions (or constant regions, CH and CL), which exhibit relatively few variations in their aminoacid sequences.

The antigen-binding sites of most antibodies are surfaces that can accommodate the conformational epitopes of different molecules. As in any molecular recognition event, the recognition of antigen by the antibody involves non-covalent reversible binding (electrostatic forces, hydrogen bonds, van der Waals forces, and hydrophobic interactions)[53]. The specific characteristics of the corresponding interaction process leads to the specific recognition of a given antigen, which may take place with varying affinities. In fact, the key recognition features of one particular antigen reflect the structural and dynamic properties of the associated antibody V regions.

1.2 Conformational analysis of carbohydrates: NMR and Molecular Dynamics

1.2.1 Anomeric and exo-anomeric effects

Carbohydrates in aqueous solution exist in equilibrium between their open chain and cyclic forms. The carbonyl carbon in the open chain becomes the anomeric carbon in the cyclized structure, which is the most oxidized carbon of the ring. Cyclized aldose and ketose carbohydrates can adopt either the α or β anomeric configuration, depending on the orientation of the group attached to the so-called anomeric carbon[54]. The assignment of the configuration is determined by the relationship between the group attached to the anomeric carbon (hydroxyl) and the group attached to the highest numbered chiral carbon in the ring (Carbon at position 6). When the groups are *trans,*

the α configuration is noted and when they are in *cis*-type orientation, the β configuration is used[55] (Fig 1.6).

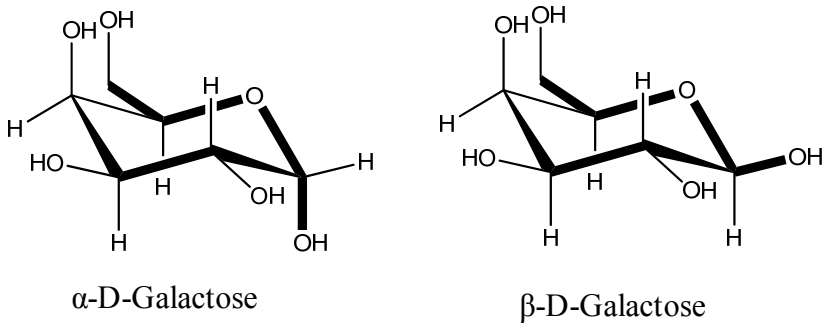


Figure 1.6 2D representation of the α and β configuration of a D-Galactose residue.

The anomeric effect was initially defined as the preference for an electronegative substituent, at the anomeric carbon in a carbohydrate, to be in an axial rather than equatorial orientation[56]. In cyclohexane, it is well known that substituents attached to the ring tend to prefer the equatorial position over the axial one, due to unfavourable steric interactions for the axial conformer. However, contrary to steric predictions, electronegative substituents at the anomeric position of carbohydrates and other heterocycles were observed to exist predominantly as the axial conformer. This phenomenon was named as the anomeric effect.

Two rationales exist that explain the origin of the anomeric effect, the electrostatic model (also known as the rabbit ear and dipole-dipole model) and the molecular orbital model (also known as the double-bond/no-bond model and hyperconjugative model)[57].

Originally, the anomeric effect was defined in terms of electrostatic interactions. In this electrostatic model,[57] it was speculated that the increased preference for the electronegative group to be axial is due to the existing repulsive dipole-dipole interactions for the equatorial orientation. However, this model was discussed by the compelling equilibrium effects that were observed when the polarity of the solvent was modified. In principle, for the electrostatic model, one would predict that a polar solvent would stabilize the more polar, equatorial, conformation to a greater extent than a non-polar solvent. Interestingly, that was not the case, evidencing that the electrostatic model is not sufficient for explaining the anomeric effect. The molecular orbital model presented an alternative explanation for the anomeric effect[56a]. The preference for the axial conformation was attributed to the stabilization provided by donation of electron

density from the anti-periplanar orbital (σ*) on the ring oxygen (sp^3 or p) to the anti-bonding orbital of the C-X bond[58]. This stabilizing interaction can only occur in the axial conformation due to the requirement for the donating oxygen orbital and the C-X bond to be anti-periplanar. The origin of the anomeric effect continues being a matter of discussion[59].

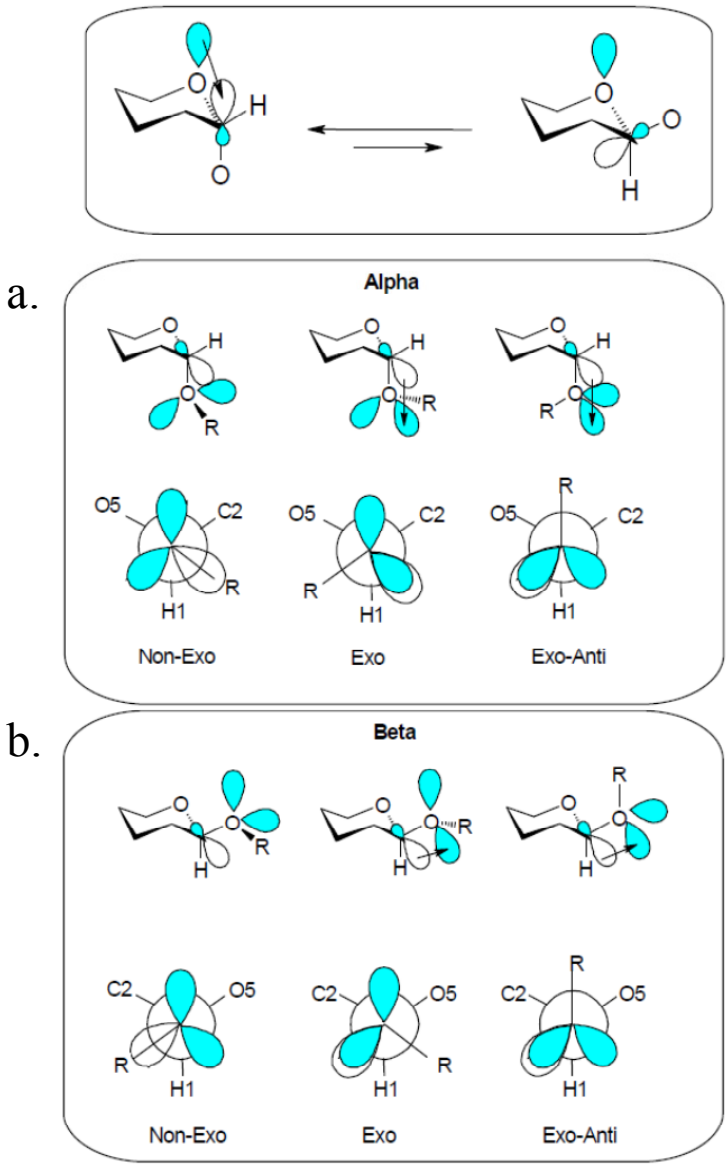


Figure 1.7 Schematic representation of the anomeric (a) and *exo*-anomeric (b) effect. {Figure Adapted from [60]}.

Similarly, the *exo*-anomeric effect describes the preference of the exocyclic glycosidic torsion angle (Φ) to adopt a *gauche* (or synclinal) orientation. In this spatial disposition, the lone pair electronic density at O1 can be delocalized with respect to the σ* from the C1-O5 bond. This fact affects the dihedral angle defined by the glycosidic

torsion and determines a favourable interaction between the glycosidic O1 electronic pair with the σ* orbital of the endocyclic bond C1-O5. This phenomenon is known as the exo-anomeric effect (out of the ring), similarly to the endo-anomeric effect (within the ring, also simply referred as the anomeric effect)[59]. In fact, the combination of steric and electrostatic interactions generates the preference for two orientations of the glycosidic torsion Φ (H1-C1-O1-Cx): -60º and +60º for the α and β forms in a $^4C_1$ chair, and +60º and -60º for the α and β forms in a $^1C_4$ chair (Fig 1.7). The magnitude of the steric and stereoelectronic contributions has been deduced in solution[60].

1.2.2 Nuclear Magnetic Resonance and Carbohydrates: Basic concepts

Nuclear Magnetic Resonance (NMR)[61] is a spectroscopic technique related with the intrinsic property of atomic nuclei that have a spin quantum number different to 0. Every nucleus have an associated magnetic moment, **μ**, by the relationship **μ = γxS**, being **γ** the magnetogyric ratio, which is characteristic of each nucleus. The magnitude and the orientation of the associated magnetic moment with the nuclear spin can only adopt a limited number of values. For a given nuclear spin n, 2n+1 possible angular moments exist, with values from –n to +n. For example, for nuclei with spin number of n=½, like ^1H, ^{15}N or ^{13}C, there are two possible spin states; m=+½ and -½.

In the absence of an external magnetic field, these levels are degenerated; that is, both states have the same energy, and the number of nuclei in both states will be about equal. On the other hand, when placed within an external, static magnetic field (denoted B_0), this degeneration disappears, producing a population difference between both states, as shown in Eq. 1.

Eq. 1

$$\Delta E = \gamma h B_0 \quad (\mathbf{1}) \qquad N_{low} / N_{app} = e^{-\Delta E / k t} \quad (\mathbf{2}) \qquad \omega_0 = \gamma (B_0 - B_i) \qquad \delta = (\omega - \omega_{ref}) / \omega_{ref}). 10^6 \quad (\mathbf{3})$$

Thus, there will be a minor fraction of nuclei populating the lower energy spin state, related with the Boltzmann distribution (Eq. 1 (2); K= Bolztmann constant).

The Nuclear Magnetic Resonance (NMR) signal results from the difference between the energy absorbed for the nuclear spins in the presence of a radiofrequency. This generates a transition from the low energy state to the high-energy state. When the

system returns to the equilibrium, the energy emitted follows the inverse step: there are transitions from the high-energy state to the low-energy state. Therefore, the NMR signal is proportional to the population difference in both states.

The magnetic moment of every nucleus precesses around the external magnetic field (B_0). The frequency of this precession is named *Larmor* resonance frequency (ω_0). This is different for each particular nucleus and is related to the energy difference between both levels. The Larmor frequency of a given nucleus depends on its magnetogyric constant γ and on the magnetic field intensity, B_0 (Eq. 1 (3)). However, there are local perturbances in the magnetic field associated with every nucleus, mainly due to the protection applied by the closer electrons.

As a result, we can use NMR as a spectroscopic technique, since the nuclei with different chemical environment precess with different frequencies. To reduce complexity in providing data, NMR prefers the use of chemical shifts (δ, expressed in ppm), to the use of actual Larmor frequencies. The corresponding relationship is shown in Eq. 1(3). Here, ω is the Larmor frequency for an observed nucleus and ω_{ref} is the Larmor frequency for a reference nucleus, which normally corresponds with TMS or TSP in ^{1}H NMR.

There has been a spectacular progress in the NMR field with advances in distinct areas to generate an improvement in the resolution (decreasing the field homogeneity), an increase in the sensibility (using the Fourier Transformation), or the development of structure determination, conformation and dynamics techniques of complex molecules.[61]

1.2.3 Conformation and NOE effect

The shape of a given oligosaccharide is determined by the conformation of the furanose/pyranose rings and the relative positioning of the rings in the chain. Based on X-Ray crystallography, neutron diffraction, and NMR data, the $^4C_1$ chair conformer is energetically preferred for D-pyranose ring structures ($^1C_4$ for L-sugars)[62]. In rare cases, for example, for L-iduronic acid pyranoses, constituents of heparan and dermatan sulfates, this is not the case. L-iduronic acid displays a major skew boat form ^2S_O[6, 63]. However, the main contribution to defining the shape of a glycan generally does not originate from this source, but rather from changes of the two dihedral angles **Φ**, **Ψ** of

each glycosidic bond (Fig. 1.8). For linkages involving the hydroxymethyl C6 carbons, a third angle, **ω**, should be also considered. Since the pyranose rings linked by the glycosidic bonds and their exocyclic substituents may be rather bulky, their sizes imposes topological restraints on the intramolecular movements of the oligomer.

The orientations of the glycosidic bonds are specified by the corresponding torsion angles **Φ/Ψ** (H1-C1-O1-Cx and C1-O1-Cx-Hx, respectively, with x being the aglyconic linkage position, as in Fig. 1.8). For 1-6 linkages, **ω** is defined as O5-C5-C6-O6.

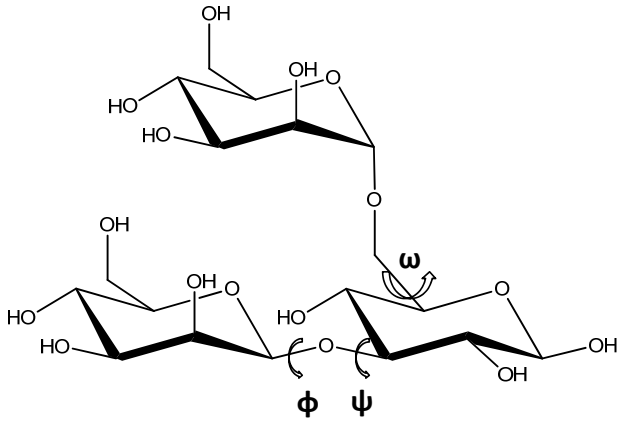


Figure 1.8 Torsion angles φ, ψ and ω.

Compared to oligopeptides with small side chains, the conformational space accessible to the molecule at room temperature is thus relatively restricted. This spatial factor limits the range of interchangeable conformations. The accessible space has been inferred by using computer-assisted molecular mechanics and dynamics calculations and convincingly documented by experimental evidence, primarily from NMR[63a, b]. Exploring the actual positions of each oligosaccharide on the scale between high flexibility with an ensemble of conformers and almost complete rigidity definitely has salient implications for predicting its role as a coding unit. In this respect, it is also worth pointing out that the existing notable level of intramolecular flexibility is not a favourable factor for crystallization. Indeed, such an extent of unrestrained conformational entropy can contribute to explaining the frequently frustrating experience in crystallization attempts in carbohydrate chemistry. If, on the other hand, the level of conformational entropy is confined to only very few stable conformers, the presented shape distribution is a function not only of the sequence, but also of external factors affecting the current status of the equilibrium. Sugar receptors as probes for

distinguishing bioactive or bioinert glycan presentation modes on proteins have already provided the hypothesis experimental support[64].

Emil Fischer already introduced the classical projection formulae for sugars, with a standard orientation (carbon chain vertical, carbonyl group at the top; Fig. 1.9).The use of models with flexible bonds between the atoms, easily permitted to 'stretch' the sugar models into a position suitable for projection. He assigned to dextrorotatory glucose (via the derived glucaric acid) the projection with the OH group at C-5 pointing to the right, well knowing that there was a 50% chance that this was wrong.

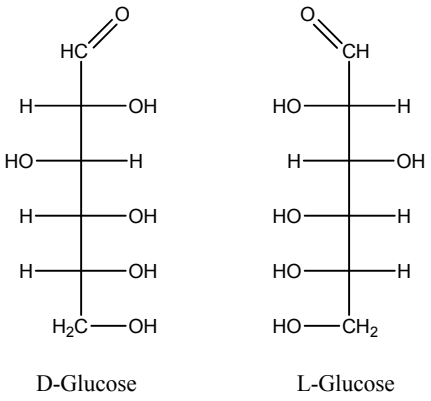


Figure 1.9 Classical representation of pyranose by Fisher.

Towards the end of the nineteenth century, it was realized that free sugars (not only the glycosides) existed as cyclic hemiacetals or hemiketals. Mutarotation, discovered in 1846 by Dubrunfaut, was now interpreted as being due to a change in the configuration of the glycosidic (anomeric) carbon atom. Fischer assumed the cyclic form to be either five-membered or six-membered rings. In the 1920s, Haworth proposed the terms 'furanose' and 'pyranose' for the two forms. He also introduced the 'Haworth depiction' for writing structural formulae, a convention that was soon widely followed. Cyclic forms with a three-membered ring are called oxiroses, those with a four-membered ring oxetoses, those with a five-membered ring furanoses, with a six-membered ring pyranoses, with a seven-membered ring septanoses, with an eight-membered ring octanoses, and so on. If a cyclic form of a sugar is to be represented in the Fischer projection, a long bond can be drawn between the oxygen involved in ring formation and the (anomeric) carbon atom to which it is linked.

In the Haworth representation of the cyclic form (Fig. 1.10), the ring is orientated almost perpendicular to the plane of the paper, but viewed from slightly

above so that the edge closer to the viewer is drawn below the more distant edge, with the oxygen behind and C-1 at the right-hand end. To define the perspective, the ring bonds closer to the viewer are often thickened. Figure 10 shows the schematic representation of pyranose ring closure in D-glucose with the reorientation at C-5 necessary to allow ring formation.

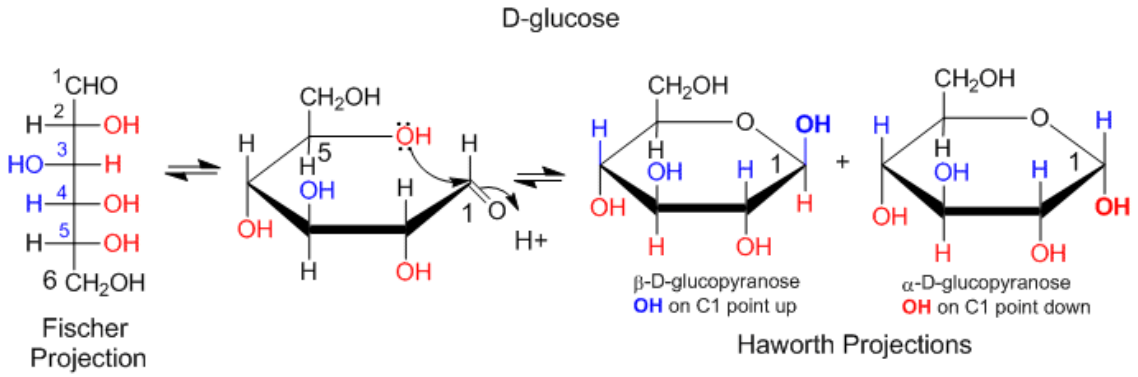


Figure 1.10 Classical representation of pyranose by Fisher and Haworth.

The orientation of the model described above results in a clockwise numbering of the ring atoms. Groups that appear to the right of the modified Fischer projection appear below the plane of the ring; those on the left appear above. In the common Haworth representation of the pyranose form of D-aldohexoses, C-6 is above the plane.

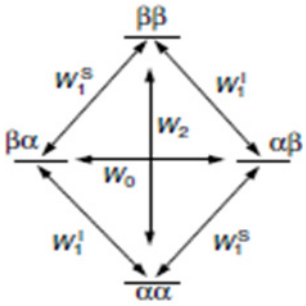
The Haworth representation implies a planar ring. However, as mentioned above, due to the tetrahedral geometric of carbon backbone, monosaccharides assume conformations that are not planar. The conformation and the spatial arrangements of the ring atoms of a monosaccharide in the cyclic form is well described by the IUPAC-IUB Joint Commission on Biochemical Nomenclature (E_2, ^3E, $^3T_2$, $^4C_1$, $^1C_4$, $^{2,5}B$, $^1S_3$, $^2S_0$, etc).

The NOE effect

The nuclear Overhauser enhancement (NOE) is the key parameter to determine internuclear (proton-proton) distances and therefore, it contains key conformational information. It is based on the existence of dipole-dipole interactions between two moving NMR-active nuclei, which must be close enough in the space.

If we consider two spins, I and S, the intensity of I changes when the equilibrium state of the closest spin S is perturbed by a radio-frequency pulse. Then, the

magnetization is transferred from one spin to the other one through space. This change is intensity is commonly known as the NOE and it is governed by three possible transitions: the zero-, single- and double quantum transitions[61, 65], which describe the cross-relaxation pathway (Fig. 1.11) and the variation of intensities, as defined by the Solomon equation given below:



W_1^S and W_1^I : single quantum transitions
W_2 : double quantum transitions
W_0 : zero quantum transitions
W_2 and W_0 are cross relaxation probabilities

Figure 1.11 Diagram illustrating the population levels, intensity changes and transition probabilities in a two-spin ½ system, which gives rise to the NOE.

Eq. 2

$$dI_Z/dt = -(I_Z - I^0_Z)(W_0 + 2W_1^I + W_2) - (S_Z - S^0_Z)(W_2 - W_0)$$

S_Z and I_Z are the longitudinal components of the magnetization. S^0_Z and I^0_Z are these values at zero time. The zero-, one and double quantum transitions depend on the tumbling rate of the molecule and determine if the variation of the intensity is negative or positive.

The cross relaxation rate constant σ_{IS} is defined as the difference between the double and zero-quantum transition probabilities (W_2-W_0) and is a measure of how fast the magnetization is transferred between the S and I spins. Finally, the dipolar longitudinal relaxation rate constant, ρ_{IS}, is defined by the term $W_0+2W_1^I+W_2$, and describes the part of the relaxation process that restore the equilibrium state of spin I.

In this way, after saturating spin S, at the steady state: $dI_Z/dt=S_Z=0$, and the equation can be rewritten as:

Eq. 3

$$(I_Z - I^0_Z) / S^0_Z = \sigma_{IS} / \rho_{IS}$$

and the enhancement of the intensity (η) of the signal of I after saturation of spin S can be determined by:

Eq. 4

$$\eta_I \{S\} = \gamma_S / \gamma_I \, [\, W_2 - W_0 / W_0 + 2W_1^I + W_2 \,] \equiv \gamma_S / \gamma_I \, [\, \delta_{IS} / \rho_{IS} \,]$$

the gyromagnetic ratios (γ_S / γ_I) are important when the two spins are different. In the case of homonuclear NOE, the σ_{IS} and ρ_{IS} are related to the precession frequency (ω) and the correlation time (τ_C):

Eq. 5

$$\eta = (5 + \omega^2\tau_C^2 - 4\omega^4\tau_C^4) / (10 + 23\omega^2\tau_C^2 + 4\omega^4\tau_C^4)$$

According to these equations, NOEs depend on the spatial proximity (for the transfer to be efficient), on the spectrometer frequency and on the tumbling rate[61, 65]. For small molecules, with fast tumbling in solution, dipolar relaxation is long and the NOE signals are positive, with a maximum enhancement of 50% (Figure 1.12). In contrast, large molecules, slowly tumbling in solution, provide efficient NOE negative signals, and the enhancements can reach 100%. This change of sign can be explained looking at the possible transitions. In fact, for short τ_C the best possible transition is the double quantum transition W_2, which corresponds to fast frequencies. When W_2 dominates, the intensity of I increases, corresponding to a NOE positive signal. The opposite behaviour is found for large molecules with large τ_C. In this case, W_0 dominates, and the intensity of I decreases, corresponding to a NOE negative signal[61, 65].

Also, for intermediate motions, the curve in Fig. 1.12 (NOE curve) passes through zero, in a region in which the enhancement of NOE is negligible. Depending on the temperature and spectrometer frequency, this zero-crossing point is found from 500 to 1500Da, (from tri- to penta-saccharides).

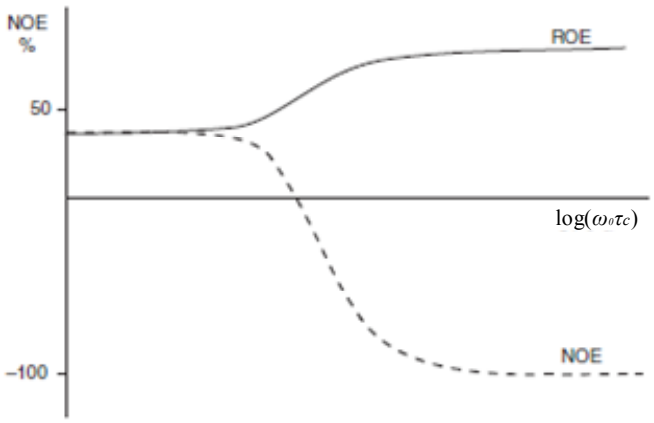


Figure 1.12 Dependence of NOE and ROE with molecular motion.

It is possible try to avoid the zero-crossing point, by modifying the experimental conditions that can affect τ_C. Spectra at lower temperatures can be recorded, or taken after modifying the solution conditions (viscosity). Also, a different spectrometer frequency can be used. In fact, it has been described that ultra-high field spectrometers (800, 900 MHz, and beyond) permit to reach a better resolution, since the spin-spin relaxation mechanism is affected. In fact, for molecules with fast τ_C, the constant rate $1/T_2$ depends on B_0 (the static external magnetic field). Therefore, at high magnetic fields, T_2 is more favorable leading to narrow signals (better resolution) and better enhancement in NOE signals[66].

In a real situation, the NOE (η) also depends on the proton-proton distances. Thus, the NOE is inversely proportional to the inverse of the sixth power of the distance between the interacting nuclei. In this case, knowing the NOE value between two nuclei (A-B) and the distance between them, it is possible to calculate other distances corresponding to observed NOEs (X and Y) in the same molecule through this simple equation (isolated spin pair approximation):

Eq. 6

$$\eta_A \{B\} / \eta_X \{Y\} = r_{AB}^{-6} / r_{XY}^{-6}$$

This distance dependence reflects the necessity for dipolar interactions (short range) to give rise to NOEs. In fact, at large distances the NOE effect is minimum, with an exponential decay in the NOE with increasing distances[65]. Depending on the molecular motion and the spectrometer frequency, the maximum distance that can be monitored in oligosaccharides is aprox. 4 Å, whereas for large molecules can reach even 6 Å.

A variety of NMR sequences have been employed to measure NOEs, both in 1D and 2D versions (Fig. 1.13). Nowadays, the so-called transient experiments are employed, instead of the saturation experiments that were commonly employed years ago (and in the Solomon equations). In this method (either 1D or 2D), the perturbation of the system is obtained by inverting the target resonance/s. Just one resonance is usually selected in the 1D version, or all the spins in the 2D-protocol. Thus, the inversion of the population differences across the corresponding transitions allows NOE development in the absence of external interference. During the mixing time, the NOE grows up for a selected period of time and then, the changes in intensities are monitored by the employment of a read pulse and acquisition[65, 67].

In all these cases, the key points are the calibration of the 90° pulse and the determination of the associated delays to obtain the best results. As mentioned above, one additional problem for oligosaccharides could be the zero crossing region around $\omega\tau_C \sim 1.1$, were the observed NOEs are basically absent. If the experimental conditions of the sample cannot be modified or the access to a different spectrometer is not possible, there is one additional way to NOEs, through the use rotating-frame NOE spectroscopy or ROESY experiment[65]. Under spin-lock conditions, the nuclei are still able to experience dipolar relaxation, now in the transverse plane, leading to ROE effects, which are always positive for all values of τ_C (Fig. 1.12, ROE curve). A frequent problem is the occurrence of scalar-coupling-mediated cross peaks, due to TOCSY-like effect during the spin-lock period. Some modification shave been employed to overcome this problems[65].

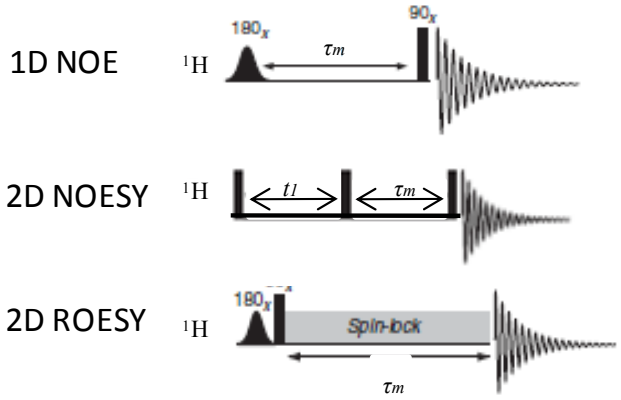


Figure 1.13 Different experimental pulse sequences to measure NOEs. 1D-NOE: A single resonance is selectively inverted. Then, NOEs grow up during mixing time (τ_m). 2D-NOESY: All resonances are frequency labeled after the 90 degree pulse and inverted after the evolution period. 1D-ROESY: NOEs in the rotating frame. All resonances are frequency labeled after the 90 degree pulse and "spin locked" after the evolution period. Rotating frame NOEs grow up during τ_m.

1.2.4 Computational Methods: Molecular Dynamics

Molecular modeling methods try to help to describe a living system through the chemical and physical points of view. It is today well known and assumed that biomolecules are dynamic entities that display their functionality by adopting distinct conformations in solution.

MD may give details on the different time-dependent properties of the system. In principle, different information may be extracted from a MD simulation:

- From the exploration of the conformational space to the refinement of a given structure.
- From the characterization of equilibrium averages to the description of the structural, motional and thermodynamic properties of the system.

For these purposes, empirical energy functions have been developed. They are mathematical equations constituted by simple terms that allow the description of physical interactions closely related to the structure and dynamic properties of biomolecules. These empirical "force fields" use atomistic models instead of electrons and nuclei to simplify the mathematical treatment and enhance the speed of calculations.

The potential energy function permits to calculate the potential energy (V) as a function of the three-dimensional structure (R), and it is divided in two main terms:

Eq. 7

$$V(R)_{total} = V(R)_{internal} + V(R)_{external}$$

$$V(R)_{internal} = \sum_{bonds} K_b(b - b_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} K_\chi[1 + \cos(n\chi - \sigma)]$$

$$V(R)_{external} = \sum_{\substack{nonbonded \\ atompairs}} \left(\varepsilon_{ij} \left[\left(\frac{R_{\min ij}}{r_{ij}} \right)^{12} - \left(\frac{R_{\min ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\varepsilon_D r_{ij}} \right)$$

The internal term gathers all the parameters related to covalently-bonded atoms, while the external one represents the non-bonding interactions between atoms[68].

The parameters characterizing a three-dimensional covalently-bonded structure (internal term) are bond lengths (b), valence angles (θ), dihedral angles (χ), and distances (r*ij*).

In any modern force-field, bond-stretching and angle-bending are treated as harmonic functions, which oscillate around an equilibrium value (b0, θ0). Kb and $K\theta$ are the force constants associated with the bond and angle terms.

When MDs are performed, in absence of bond-breaking or bond-making events, motions are strongly related to the stretching, bending and torsions of the connected atoms. The torsional term is treated as an oscillatory function. In fact, in addition to the force constant $K\chi$, it also includes periodicity (n) and phase (δ) terms. The magnitude of $K\chi$ is the height of the energy barrier to rotation, n indicates the number of cycles per 360° rotation, while δ is the location of the maxima in the dihedral energy surface, generally 0º or 180º. Each torsion angle in a molecule is treated as a sum of dihedral terms that have different periodicity, constant forces and phases[68].

The external terms are of paramount importance to characterize biological systems. There is also a strong influence of the environment on the properties of biomacromolecules. The key parameters of the external terms are the well depth ε_{ij} between the atoms i and j, the minimum interaction radius, R_{minij}; and the partial atomic charge, q. The term in square brackets is associated to the van der Waals (VDW)

interactions. In many cases, a Lennard-Jones potential (or 6-12 potential) is used to define these interactions (Fig 1.14). The $(1/r)^{12}$ represents the repulsion energy between atoms, very important at short-range distances. The attraction force energy is defined by the $(1/r)^6$ term and is related to instantaneous dipole-induced dipole interactions (London's dispersion forces). This term works at longer distances. The negative sign denotes the favourable nature of this phenomenon[68]. The minimum interaction radius R_{minij} and the well depth, ε_{ij} (Fig. 1.14)[69] are related to the minimum of the curve. They are not determined for every possible interaction pair, but calculated for the individual atom types and are combined to obtain the *ij* cross terms.

Eq. 8

$$R_{minij} = (R_{mini} + R_{min\,j}) / 2$$

Eq. 9

$$\varepsilon_{ij} = \sqrt{(\varepsilon_i \varepsilon_j)}$$

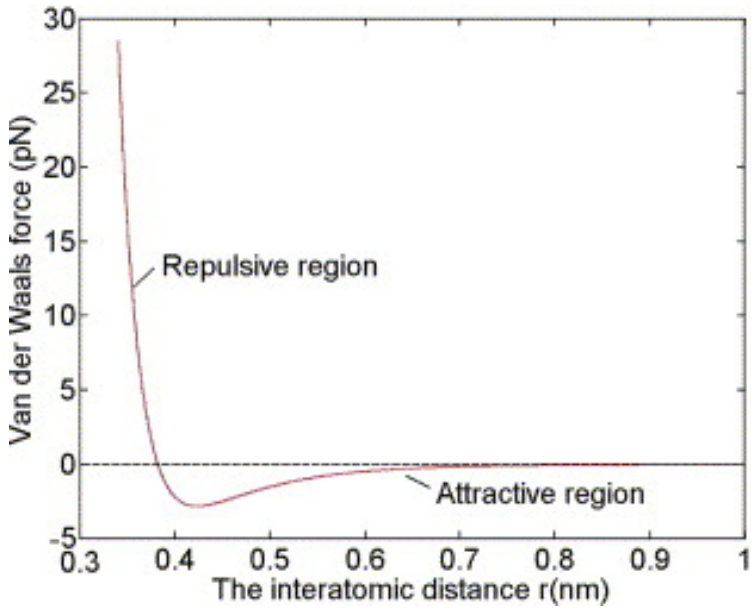


Figure 1.14 Variation of the van der Waals energy versus distance.

The last term of the external contribution to the total potential energy is the electrostatic or Coulomb term. It defines the distance-dependent interactions between the "partial" charges q*i* and q*j* related to atoms *i* and *j*.

The non-bonded terms (VDW and electrostatic) are essential to describe the conformation of molecules and their interactions with the environment. They strongly influence the calculation, since there are many non-bonded contacts, many more than bond, angle and torsion interactions[68].

The non-bonded dispersion and repulsion terms rapidly decay with the distance. Therefore, the calculations can be restricted to the estimations between neighbouring nuclei. Generally, a cut-off is defined, around 8-10 Å[65, 68].

There are several force fields (FFs) for biomolecules[70]. Some are well parametrized for proteins, peptides, carbohydrates, and small organic molecules, or combinations thereof. Some may include explicit solvent representations, or implicit solvent models, using a dielectric constant to simulate the medium[71]. Others may estimate free energies. It is important to select the proper set of parameters (FFs) related to the particular application, especially if the calculated results are to be compared with experimental data. From this perspective, it is also essential to define the required experimental information. We have used the TIP3P rigid water model, which considers the three atoms of water, each atom with a point charge, and Lennard-Jones parameters applied on the oxygen atom[65]. Due to its simplicity and efficiency, this is the most widely used in MDs. When calculations with explicit solvent are performed, the solute is placed in a box under so-called "periodic boundary conditions". The solute and the solvent are arranged in a regular lattice structure, and a unit cell is generated, so that the atoms (and their associated properties) are infinitely replicated in three dimensions. In this way, it is possible to avoid edge effects. In fact, when a molecule goes out from one side of the box, it re-enters from the other side. This method permits to maintain a constant density and, depending on the protocol, also the internal pressure can be constant. The shape of the box can be cubic (See Fig. 1.15), a truncated octahedron, or a rhombic dodecahedron.

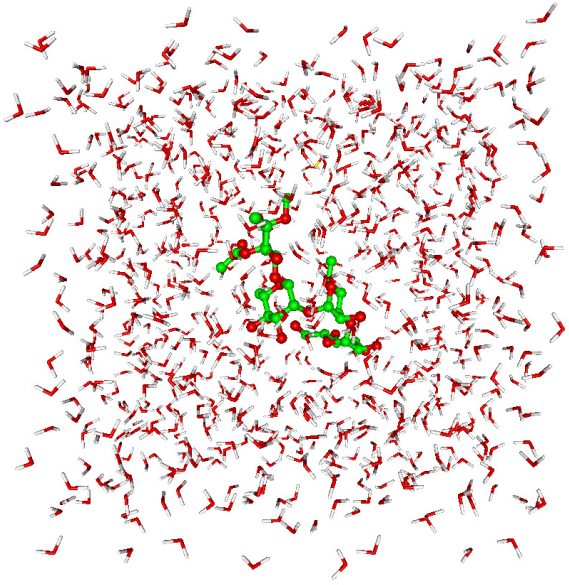


Figure 1.15 Approach for solvated MD calculation; a cubic box.

In this Thesis, AMBER11 has been used for all the MD simulations. The AMBER software package is an ensemble of programs[72] that allow different type of calculations, including MDs, free energy calculations, QM/MM simulations, MDs with time-average restrains from experimental data, etc. Several force fields are implemented for the different biomolecules, Parm99[73] and its variations for proteins and nucleic acids[74], Glycam06 for carbohydrates[75], and GAFF, which is the general AMBER force field[76], for organic molecules.

The following programs, presented in Fig. 1.16, are used to prepare the input files, to perform the simulations, and to analyze the data:

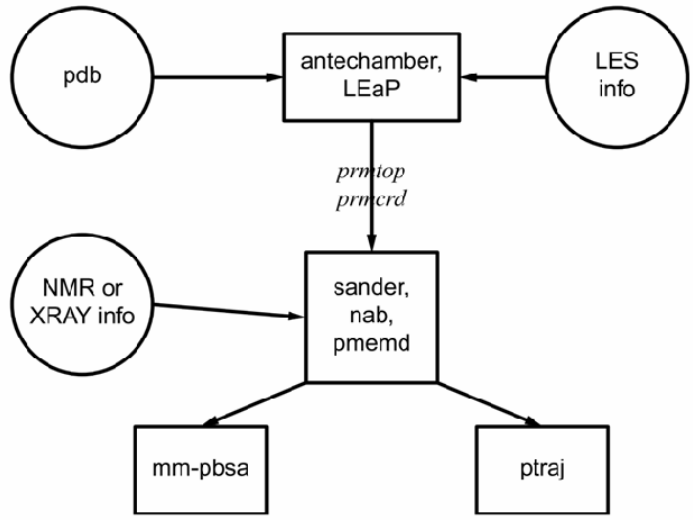


Figure 1.16 Basic information flow in AMBER[77]

1.3 NMR and Molecular Recognition

NMR spectroscopy is one of the most widely used techniques to characterize molecular recognition events at the atomic level. NMR techniques can be employed for studying protein-ligand interaction process. These methods, usually assisted by computational protocols, permit to elucidate the structural features of the binding entities, as well as to characterize the kinetic and thermodynamic characteristics of the free-bound equilibrium. NMR techniques provide the possibility to study interaction events directly in solution, close to the physiological medium. Since the binding process strongly modifies the physicochemical properties of both partners, especially rotational and diffusion motion timescales, there are several NMR parameters that are influenced by the interaction process, as chemical shift, relaxation times and NOEs, diffusion constants, etc.

The binding of small ligands to large receptors follows a bimolecular association reaction with second-order kinetics:

Eq. 10

$$[E] + [L] \underset{k_{off}}{\overset{k_{on}}{\leftrightarrow}} [EL]$$

The equation represents a dynamic equilibrium involving three species: the free receptor E, the free ligand L, and the receptor-ligand complex, EL. The unimolecular rate constant k_{off} is inversely proportional to the mean lifetime τB of the receptor-ligand complex. The bimolecular rate constant, k_{on}, measures the probability of a productive encounter between the free receptor and the ligand and is often assumed to be controlled by diffusion. Thus, k_{on} can vary between 10^7 and 10^9 M^{-1} s^{-1}. The binding affinity can be described by the temperature-dependent equilibrium dissociation constant or K_D:

Eq. 11

$$K_D = \frac{[E] \cdot [L]}{[EL]} = \frac{k_{off}}{k_{on}}$$

The bound receptor fraction is defined by

Eq. 12

$$P_B^E = \frac{[EL]}{[E]+[EL]}$$

Last equations can be combined yielding the following equation:

Eq. 13

$$P_B^E = \frac{[L]}{[L]+K_D}$$

where $P_B{}^E$ is the fractional occupation of the receptor binding site by ligand L.

When [L]«K_D, $P_B{}^E$ is proportional to [L]. When [L]=K_D, the receptor is half saturated; that is, half of the receptor molecules exist in a one-to one complex with the ligand. When [L]»K_D, the receptor is completely saturated and $P_B{}^E$= 1; in this limit, every receptor binding site is occupied by one ligand, which on average exchanges with a second distinct ligand every ≈ $1/k_{off}$ seconds. According to these points, ligands with weak affinity have large K_D and thus require the addition of more ligand to saturate the receptor binding site.

In the two-state equilibrium, ligand and receptor molecules will exist in either a free (L, E) or complexed (EL) state. In the free state, both receptor and ligand retain their intrinsic NMR parameters (e.g. chemical shifts, relaxation rates, translational diffusion coefficients). In each other's presence, the mutual binding affinity of ligand and receptor drives an exchange process that toggles both sets of molecules between the free and complexed states. At equilibrium, they adopt free and bound state populations ([E], [L], [EL]). Under these conditions, the ligand transiently adopts NMR parameters characteristic of the typically much larger receptor. Alternatively, from the receptor's perspective, the ligand transiently perturbs the binding site microenvironment(s), which may alter the distribution of conformations sampled by the ensemble of receptor molecules. In either case, the exchange modulates the NMR parameters of both molecules. It is the ability to experimentally distinguish these exchange-modulated parameters from those of the free state what enables NMR detection of the receptor-ligand interaction and hence the employment of NMR-based screening protocols.

A complete discussion of how chemical exchange modifies the NMR parameters (dynamic NMR) should include a description of the modified Bloch equation formalism of Hahn, Maxwell, and McConnell[78]. This formalism (HMM) provides an excellent theoretical framework to describe the majority of the exchange phenomena that occur in NMR screening experiments, under all exchange regimes. For most applications described here, it is sufficient to consider only the case of fast exchange.

1.3.1 The thermodynamics and kinetic aspects

The solutions to the HMM equations[78] describe the behaviour of the system magnetization on arbitrary exchange time-scales. In NMR-based screening, however, these equations are almost never solved, and fast exchange is simply assumed. This assumption is made for two reasons. First, the experimental conditions for ligand-based NMR studies are often well suited to fast exchange. These experiments are typically carried out with $L_T/E_T > 10$ (where L_T and E_T are total ligand and total receptor concentration respectively) and the binding compounds have a $K_D \geq 100\mu M$. If k_{on} is well approximated by a diffusion-limited value (10^7-10^9 M^{-1} s^{-1}), then the slowest exchange rate k_{ex} value will lie in the range $1000 < k_{ex} < 100\,000$ s^{-1}. Ligand-based NMR methods are primarily ^{1}H-based and consequently, k_{ex} exceeds most differences in intrinsic ^{1}H relaxation rates and rotating frame precession frequencies. This fact thus ensures that the fast exchange assumption is valid. A second motivation for assuming fast exchange is the resulting algebraic simplicity. Generally, the NMR parameters Q become the simple averages.

Eq. 14

$$Q_{avg} = P_B Q_B + P_F Q_F$$

Here, Q_{avg} is the observed exchange-averaged parameter for the ligand (or receptor) in the presence of the receptor (or ligand). Observed differences between Q_{avg} and Q_F provide a signature of receptor binding and indicate a hit in a NMR screening based on that parameter. In the case of Eq.14, Q_{avg} is a simple population-weighted average and it applies to those parameters Q for which chemical shift modulations are not relevant. These parameters include longitudinal auto-relaxation and cross-relaxation

rates, rotating-frame spin-locking autorelaxation and cross-relaxation rates, as well as translational diffusion coefficients. The bound state contribution in Eq.14 is $P_B Q_B$. The ability to detect binding with adequate sensitivity depends critically on $P_B Q_B$, being significant relative to $P_F Q_F$. However, typical screening conditions where $L_T \gg E_T$ make $P_B \ll P_F$. For this reason, it is much preferred to measure NMR parameters Q that become amplified in the bound state (i.e. $Q_B \gg Q_F$).

1.3.2 The receptor point of view

Screening by NMR may proceed by using ligand- or receptor-based methods. Receptor-based methods observe and compare the NMR parameters of the receptor molecule resonances in the presence and absence of compound mixtures. Thus far, the receptor-based methods have focused mostly on proteins. Such methods incorporate site-specific characterization afforded by assigned protein NMR spectra along with *a priori* knowledge of the protein's three-dimensional structure (either from X-ray, NMR, or modeling procedures) to drive lead generation. By identifying perturbations of assigned protein resonances, not only are ligands identified, but also their binding sites are localized. Localization of binding sites also enables one to immediately distinguish specific from nonspecific binding. Finally, unlike ligand-based methods, receptor-based methods do not rely on fast exchange to retrieve bound state information. Observation of receptor resonances permits the characterization of both higher and lower affinity hits.

NMR methods demand physicochemical properties of the protein target that present progressively more difficult challenges. First, milligram quantities of soluble, non-aggregated protein must be over-expressed and purified. Then, suitable expression hosts must be found that permit isotope enrichment (e.g. ^{13}C, ^{15}N, ^{2}H) critical for the resonance assignment of typically large (>30000Da) therapeutic targets. After sufficient quantities of labelled protein are available, it must be ensured that the sample is stable for the time required for sequential resonance assignment. Although new data acquisition approaches promise to accelerate resonance assignment, it can still be a relatively lengthy process (weeks) for the large monomeric proteins (>30000Da) routinely encountered in pharmaceutical research. Unfortunately, the time required for NMR assignment of such targets inevitably favours other approaches, such as X-ray crystallography, that can provide high-resolution structural information to medicinal

chemistry on a faster time scale (if crystals are available). The main research summarized in this Thesis refers to ligand-based NMR methods and in particular to trNOE and STD techniques. Next sections will explain, in detail, the fundamental of these methodologies.

1.3.3 The ligand point of view: trNOE and STD

Alternatively, the typical implementation of ligand-based methods compares the NMR parameters of a mixture of compounds in the presence and absence of the receptor molecules. This approach renders the molecular weight of the receptor molecule irrelevant. In fact, the most powerful ligand-based approaches become more sensitive when dealing with larger receptors. Additionally, ligand observation bypasses the need to produce milligram quantities of isotope labeled receptor. Depending on the approach, less than a milligram of unlabeled protein is required for these experiments (receptor concentration is often $\leq$1μM, and no assignments are necessary). Both receptor- and ligand-based approaches have distinct advantages and disadvantages. An obvious disadvantage of ligand-based approaches is the inability to localize the binding site of the small molecule hits on the receptor. Also, ligand-based approaches rely on the exchange-mediated transfer of bound state information to the free state. This requirement biases ligand-based methods towards the identification of weakly binding ligands (rapid exchange) and the use of large ligand molar excesses (L_T/E_T»1). The consequent risk is that, under these conditions, ligand may start to occupy, weaker affinity, non specific binding sites.

STD[79], transferred NOE (TR-NOESY), and DOSY[80] are very useful techniques detecting binding based on changes in the rotational and translational motional properties of the oligosaccharides upon binding to large macromolecular receptors[81].

- *TrNOESY*

The exchange-transferred (trNOESY) experiment is an useful technique to study the carbohydrate conformation in the presence of a specific protein. The reasons for its success probably involve various factors: sugar interactions in solution are not strong, there is fast exchange between the free and bound states of the saccharide, and the

possible changes in the conformational equilibrium of the oligosaccharide ligand upon binding to the receptor are accessible to TR-NOESY analysis.

These experiments are specifically useful in the medium–low-affinity range and they have therefore been adopted to detect carbohydrate interactions in different systems[81c]. Moreover, when a protein cannot be completely analyzed in solution, due to its large size or to the unavailability of isotopically labelled material, information on the conformation of the bound oligosaccharide can still be obtained from transferred NOE (TR-NOE) experiments. Indeed, it is well established that NOE effects (NOEs) are extremely useful in determining the 3D structure of molecules in solution[79]. When ligand molecules bind to receptor proteins, their associated NOEs undergo drastic changes in sign and magnitude due to the change in the effective global motion rotational correlation time (see below). These changes are the basis for a variety of experimental schemes that are designed to detect and characterize binding activity.

The principles of trNOEs were originally observed and described long ago and have found widespread use in the determination of the 3D structures of ligands bound to receptor proteins. The observation of trNOEs relies on the existence of different tumbling times τ_c of free and bound molecules. Low- or medium-molecular-weight molecules (M_W<2000) have a short correlation time τ_c (picoseconds) and, as a consequence, such molecules exhibit positive NOEs (or no NOEs), depending on their molecular weight, shape, and the field strength. Large molecules, however, exhibit strong and negative NOEs. When a small molecule or ligand is bound to a large molecular weight protein, it behaves as a part of the large molecule and adopts the corresponding NOE behaviour, that is, it shows strong negative NOEs (called trNOEs). These trNOEs contain information on the bound conformation of the ligand (Fig. 1.17).

Therefore, the existence of binding of a given ligand to a receptor protein can easily be distinguished by merely observing the sign and size of the observed NOEs. As for regular NOEs, various experimental implementations have been explored in the last two decades, ranging from 1D to 2D transient NOE experiments. Also, several schemes have been developed that allow a quantitative interpretation of trNOEs, thus, yielding more reliable information on the conformation of the bound ligands.

In general, it is possible to observe inter- and intramolecular trNOEs. Whereas intramolecular trNOEs are the key to define bound-ligand conformations, intermolecular trNOEs occur between a ligand and a receptor protein, and therefore, in

principle, may allow the determination of the orientation of bound ligands in protein binding pockets. From a technical perspective, it is essential to systematically employ different mixing times and protein/ligand molar ratios (usually from 5:1 to 50:1, depending on the affinity and kinetic parameters) in order to deduce quantitative conclusions, which can be further assessed by employing available computer programs[81c].

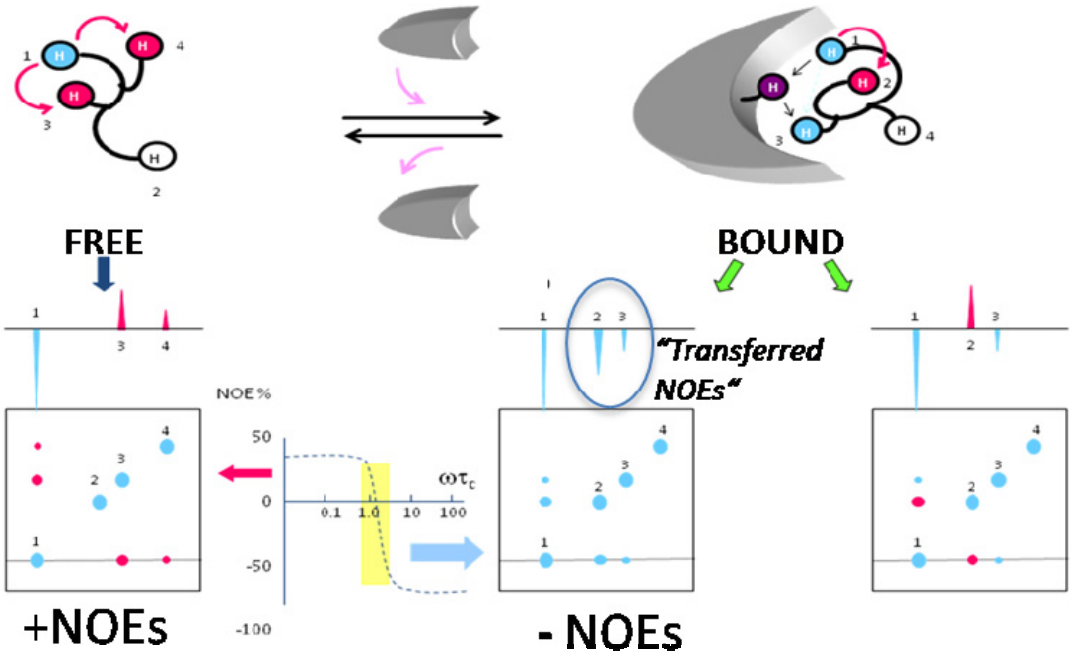


Figure 1.17 Left. Schematic representation of a NOESY spectrum for a free sugar. Cross peaks and diagonal peaks have different signs. Right. Schematic representation of a TR-NOESY spectrum recorded for an exchanging ligand–protein system. Cross peaks and diagonal peaks have the same signs, as expected for a large molecule, thus indicating binding to the protein. The relative sizes of the peaks and the appearance of new ones may be used to detect conformational variations[81c].

- *STD*

Saturation-transfer NMR spectroscopy has been used for many years to characterize binding features of ligands and receptors[82]. More recently, Meyer and Peters have developed a method based on the transfer of saturation from the protein to the bound ligands[79]. By utilizing difference spectroscopy, especially ^{1}H-based NMR experiments, it is possible to obtain well-resolved spectra of the ligand with the corresponding information. By subtracting one spectrum in which the protein

resonances are saturated from a second one, without protein saturation, a difference spectrum (see below for further details) is produced in which only the signals of the ligand(s) that have felt the transfer of saturation from the protein do appear (Fig. 1.18).

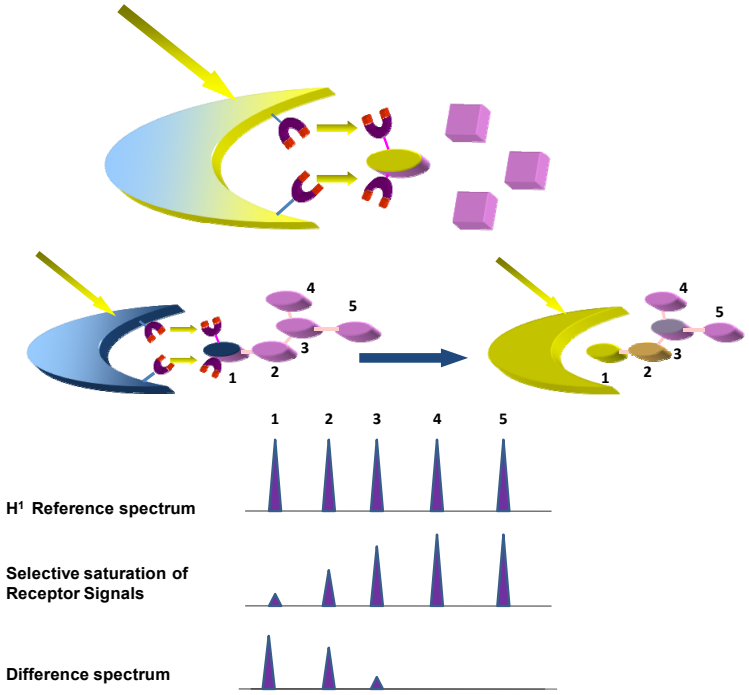


Figure 1.18 Schematic representation of the STD process[82].

The irradiation frequency is set at a position where only resonances from the protein nuclei and no resonances from the ligand nuclei are located. Therefore, in the on-resonance experiment, selective saturation of the signals of the protein nuclei is achieved. For the on-resonance irradiation, frequency values around -1 ppm are practical, because no ligand nuclei resonances are found in this spectral region, whereas the significant line width of protein signals still allows selective saturation. If the ligands show no resonance signals in the aromatic proton spectral region, the saturation frequency may also be placed there (ca. 7 ppm) or even further downfield (δ=10–12 ppm). In order to achieve the desired selectivity and to avoid side-band irradiation, shaped pulses are employed for the saturation of the protein signals[79].

The degree of ligand saturation depends on the residence time of the ligand in the protein-binding pocket. The dissociation of the ligand will then transfer the saturated ligand into solution, where the free ligand gives rise to resonance signals with narrow line widths. For those ligand protons that interact with protein protons through an

intermolecular NOE, a decrease in intensity is observed. However, in the presence of other molecules such as impurities and other non-binding components it is not usually possible to identify such attenuated signals. Therefore, in a second experiment, the irradiation frequency is set at a value that is far from any signal, ligand or protein, for example, δ40 ppm (off-resonance spectrum). The spectrum is recorded and yields the regular NMR spectrum of the mixture. Subtraction of the on-resonance from the off-resonance spectra leads to a difference spectrum, in which only the signals of the protons that were attenuated by saturation transfer are visible. All molecules without binding activity are cancelled out.

Saturation of the protein and the bound ligand is very fast (about 100 ms). Therefore, a fast off-rate of the ligand transfers the information about saturation quickly into solution. If a large excess of the ligand is present, one binding site can be used to saturate many ligand molecules in a few seconds. Ligands in solution lose their information by normal T_1/T_2 relaxation, which is in the order of seconds for small molecules. Thus, the proportion of saturated ligands in solution increases during the saturation time, and so the information about the bound state resulting from the saturated protein is amplified, which means that only a relatively small amount of protein is required. On the other hand, if binding is very tight, and consequentially off rates are in the range of 0.1–0.01 Hz, the saturation transfer to ligand molecules is not efficient. This is usually the case for K_D values less than 1 nM. If the K_D values are 100 nM or larger, fast exchange of free and bound ligands leads to an efficient buildup of saturation of the ligand molecules in solution.

The observed intensities of the signals arising from the ligand in the STD NMR spectrum are not directly proportional to the binding strength. STD NMR effects largely depend on the off-rate. As outlined above, larger off-rates should result in larger STD signals. However, when binding becomes very weak, the probability of the ligand being in the receptor site becomes very low, which results in weak STD signals.

The intensity of the STD signals depends, among other things, on the irradiation time/saturation time and on the excess of ligand molecules used[79]. The more ligand that is used and the longer the irradiation time, the stronger the STD signal is. In general, the irradiation time is 2s and a 100-fold excess of ligand is used to give good results. Before ligands in solution have lost their saturation, the process of association (followed by dissociation) can occur many times and thus put many more saturated ligands into

solution. The maximum net effect of saturation on ligand protons occurs if a large excess of ligand is used, because it is very unlikely that a ligand with saturation re-enters the binding site. From the high ligand:protein ratios, it is clear that the amount of protein required for the measurements is very small. At 500 MHz, an amount of approximately 1.2 nmol of protein is sufficient to record good STD spectra. At a molecular weight of 50 kDa, this translates into about 60 μg of protein.

It has been already shown that STD-NMR is an excellent technique for determining the binding epitope of the ligand, information that is of primarily importance for a rational drug design process[83]. As consequence of the intermolecular saturation transfer process described above, the ligand saturation is higher for those protons that are in closer contact with the receptor. This kind of information can be accurately analyzed[84] and the STD effect can be quantitatively analyzed in a similar way to that mentioned above for quantitative estimation of trNOEs[79].

The STD principle can be combined with any NMR pulse sequence, and one rather powerful experiment is the STD TOCSY experiment. Especially in cases where the library or ligand is more complex (for example, large oligosaccharides) the additional deconvolution of signals brought about by the second dimension is very helpful. The STD TOCSY experiment was also shown to be ideally suited for mapping the binding epitope of ligands[79]. By using the STD TOCSY method it was possible to screen a library of like carbohydrates to identify functional groups involved in binding[85].

STD NMR spectroscopy is well-suited to study the interactions of ligands with macromolecular receptors. As STD is a preparatory pulse sequence all types of NMR spectra can be recorded, for example, 1D NMR spectra, TOCSY, COSY, and HSQC. STD NMR spectroscopy can be used to identify from large complex mixtures of compounds those with bioactivity. Other experiments have combined the STD sequence with editing/filtering schemes based on the existence of ^{19}F, ^{13}C or ^{15}N in the interacting molecules[86].

1.4 References

1. (a) Varki, A., Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **1993,** *3* (2), 97-130; (b) Dwek, R. A., Glycobiology: Toward Understanding the Function of Sugars. *Chem Rev* **1996,** *96* (2), 683-720; (c) Gabius, H. J.; Siebert, H. C.; Andre, S.; Jimenez-Barbero, J.; Rudiger, H., Chemical biology of the sugar code. *Chembiochem* **2004,** *5* (6), 740-64; (d) Solis, D.; Jimenez-Barbero, J.; Kaltner, H.; Romero, A.; Siebert, H. C.; von der Lieth, C. W.; Gabius, H. J., Towards defining the role of glycans as hardware in information storage and transfer: basic principles, experimental approaches and recent progress. *Cells Tissues Organs* **2001,** *168* (1-2), 5-23.
2. Smith, A. E.; Helenius, A., How viruses enter animal cells. *Science* **2004,** *304* (5668), 237-42.
3. Anthony, R. M.; Ravetch, J. V., A novel role for the IgG Fc glycan: the anti-inflammatory activity of sialylated IgG Fcs. *J Clin Immunol* **2010,** *30 Suppl 1*, S9-14.
4. Gabius, H., The sugar Code: Fundamentals of Glycosciences. *Wiley-VCH* **2009,** *28*.
5. (a) Enriquez-Navas, P. M.; Marradi, M.; Padro, D.; Angulo, J.; Penades, S., A solution NMR study of the interactions of oligomannosides and the anti-HIV-1 2G12 antibody reveals distinct binding modes for branched ligands. *Chemistry* **2011,** *17* (5), 1547-60; (b) Silipo, A.; Larsbrink, J.; Marchetti, R.; Lanzetta, R.; Brumer, H.; Molinaro, A., NMR spectroscopic analysis reveals extensive binding interactions of complex xyloglucan oligosaccharides with the Cellvibrio japonicus glycoside hydrolase family 31 alpha-xylosidase. *Chemistry* **2012,** *18* (42), 13395-404; (c) Silipo, A.; Sturiale, L.; De Castro, C.; Lanzetta, R.; Parrilli, M.; Garozzo, D.; Molinaro, A., Structure of the lipopolysaccharide isolated from the novel species Uruburuella suis. *Carbohydr Res* **2012,** *357*, 75-82.
6. Nieto, L.; Canales, A.; Gimenez-Gallego, G.; Nieto, P. M.; Jimenez-Barbero, J., Conformational selection of the AGA*IA(M) heparin pentasaccharide when bound to the fibroblast growth factor receptor. *Chemistry* **2011,** *17* (40), 11204-9.
7. Perepelov, A. V.; Shekht, M. E.; Liu, B.; Shevelev, S. D.; Ledov, V. A.; Senchenkova, S. N.; L'Vov V, L.; Shashkov, A. S.; Feng, L.; Aparin, P. G.; Wang, L.; Knirel, Y. A., Shigella flexneri O-antigens revisited: final elucidation of the O-acetylation profiles and a survey of the O-antigen structure diversity. *FEMS Immunol Med Microbiol* **2012,** *66* (2), 201-10.
8. Toone, E., Structure and energetics of protein-carbohydrate complexes. *Current Opinion in Structural Biology* **1994,** *4*, 719-728.
9. Weis, W. I.; Drickamer, K., Structural basis of lectin-carbohydrate recognition. *Annu Rev Biochem* **1996,** *65*, 441-73.
10. Lobsanov, Y. D.; Gitt, M. A.; Leffler, H.; Barondes, S. H.; Rini, J. M., X-ray crystal structure of the human dimeric S-Lac lectin, L-14-II, in complex with lactose at 2.9-A resolution. *J Biol Chem* **1993,** *268* (36), 27034-8.
11. Lumry, R.; Rajender, S., Enthalpy-entropy compensation phenomena in water solutions of proteins and small molecules: a ubiquitous property of water. *Biopolymers* **1970,** *9* (10), 1125-227.
12. Vyas, N. K., Atomic features of protein-carbohydrate interactions. *Current Opinion in Structural Biology* **1991,** *1*, 732-740.

13. Hindsgaul, O., Molecular recognition. III. The binding of the H-type 2 human blood group determinant by the lectin I of *Ulex europaeus*. *Can J Chem* **1985,** *63*, 2653-2658.
14. (a) Lemieux, R., The origin of the specificity in the recognition of oligosaccharides by proteins. *Chem Soc Rev* **1989,** *18*, 347-374; (b) Lemieux, R. U., How water provides the impetus for molecular recognition in aqueous solution. *Acc. Chem. Res.* **1996,** *46* (29), 373–380.
15. Carver, J. P.; Michnick, S. W.; Imberty, A., Oligosaccharide-protein interactions: a three-dimensional view. *Ciba Found Symp* **1989,** *145*, 18-26.
16. Elgavish, S.; Shaanan, B., Structures of the Erythrina corallodendron lectin and of its complexes with mono- and disaccharides. *J Mol Biol* **1998,** *277* (4), 917-32.
17. Chavez, M. I.; Andreu, C.; Vidal, P.; Aboitiz, N.; Freire, F.; Groves, P.; Asensio, J. L.; Asensio, G.; Muraki, M.; Canada, F. J.; Jimenez-Barbero, J., On the importance of carbohydrate-aromatic interactions for the molecular recognition of oligosaccharides by proteins: NMR studies of the structure and binding affinity of AcAMP2-like peptides with non-natural naphthyl and fluoroaromatic residues. *Chemistry* **2005,** *11* (23), 7060-74.
18. (a) Jimenez-Barbero, J.; Javier Canada, F.; Asensio, J. L.; Aboitiz, N.; Vidal, P.; Canales, A.; Groves, P.; Gabius, H. J.; Siebert, H. C., Hevein domains: an attractive model to study carbohydrate-protein interactions at atomic resolution. *Adv Carbohydr Chem Biochem* **2006,** *60*, 303-54; (b) Asensio, J. L.; Arda, A.; Canada, F. J.; Jimenez-Barbero, J., Carbohydrate-aromatic interactions. *Acc Chem Res* **2013,** *46* (4), 946-54.
19. (a) del Carmen Fernandez-Alonso, M.; Canada, F. J.; Jimenez-Barbero, J.; Cuevas, G., Molecular recognition of saccharides by proteins. Insights on the origin of the carbohydrate-aromatic interactions. *J Am Chem Soc* **2005,** *127* (20), 7379-86; (b) Vandenbussche, S.; Diaz, D.; Fernandez-Alonso, M. C.; Pan, W.; Vincent, S. P.; Cuevas, G.; Canada, F. J.; Jimenez-Barbero, J.; Bartik, K., Aromatic-carbohydrate interactions: an NMR and computational study of model systems. *Chemistry* **2008,** *14* (25), 7570-8.
20. Zapico, J. M.; Serra, P.; Garcia-Sanmartin, J.; Filipiak, K.; Carbajo, R. J.; Schott, A. K.; Pineda-Lucena, A.; Martinez, A.; Martin-Santamaria, S.; de Pascual-Teresa, B.; Ramos, A., Potent "clicked" MMP2 inhibitors: synthesis, molecular modeling and biological exploration. *Org Biomol Chem* **2011,** *9* (12), 4587-99.
21. (a) Rabinovich, G. A.; Toscano, M. A.; Jackson, S. S.; Vasta, G. R., Functions of cell surface galectin-glycoprotein lattices. *Curr Opin Struct Biol* **2007,** *17* (5), 513-20; (b) Minko, T., Drug targeting to the colon with lectins and neoglycoconjugates. *Adv Drug Deliv Rev* **2004,** *56* (4), 491-509.
22. Ghazarian, H.; Idoni, B.; Oppenheimer, S. B., A glycobiology review: carbohydrates, lectins and implications in cancer therapeutics. *Acta Histochem* **2011,** *113* (3), 236-47.
23. Bouckaert, J.; Berglund, J.; Schembri, M.; De Genst, E.; Cools, L.; Wuhrer, M.; Hung, C. S.; Pinkner, J.; Slattegard, R.; Zavialov, A.; Choudhury, D.; Langermann, S.; Hultgren, S. J.; Wyns, L.; Klemm, P.; Oscarson, S.; Knight, S. D.; De Greve, H., Receptor binding studies disclose a novel class of high-affinity inhibitors of the Escherichia coli FimH adhesin. *Mol Microbiol* **2005,** *55* (2), 441-55.
24. (a) Asensio, J. L.; Canada, F. J.; Siebert, H. C.; Laynez, J.; Poveda, A.; Nieto, P. M.; Soedjanaamadja, U. M.; Gabius, H. J.; Jimenez-Barbero, J., Structural basis for chitin recognition by defense proteins: GlcNAc residues are bound in a multivalent fashion by extended binding sites in hevein domains. *Chem Biol* **2000,** *7* (7), 529-43; (b)

Asensio, J. L.; Siebert, H. C.; von Der Lieth, C. W.; Laynez, J.; Bruix, M.; Soedjanaamadja, U. M.; Beintema, J. J.; Canada, F. J.; Gabius, H. J.; Jimenez-Barbero, J., NMR investigations of protein-carbohydrate interactions: studies on the relevance of Trp/Tyr variations in lectin binding sites as deduced from titration microcalorimetry and NMR studies on hevein domains. Determination of the NMR structure of the complex between pseudohevein and N,N',N''-triacetylchitotriose. *Proteins* **2000,** *40* (2), 218-36; (c) Asensio, J. L.; Canada, F. J.; Bruix, M.; Rodriguez-Romero, A.; Jimenez-Barbero, J., The interaction of hevein with N-acetylglucosamine-containing oligosaccharides. Solution structure of hevein complexed to chitobiose. *Eur J Biochem* **1995,** *230* (2), 621-33.

25. Hernandez-Gay, J. J.; Arda, A.; Eller, S.; Mezzato, S.; Leeflang, B. R.; Unverzagt, C.; Canada, F. J.; Jimenez-Barbero, J., Insights into the dynamics and molecular recognition features of glycopeptides by protein receptors: the 3D solution structure of hevein bound to the trisaccharide core of N-glycoproteins. *Chemistry* **2010,** *16* (35), 10715-26.

26. Beintema, J. J., Structural features of plant chitinases and chitin-binding proteins. *FEBS Lett* **1994,** *350* (2-3), 159-63.

27. Rodriguez-Romero, A.; Ravichandran, K. G.; Soriano-Garcia, M., Crystal structure of hevein at 2.8 A resolution. *FEBS Lett* **1991,** *291* (2), 307-9.

28. Wagner, S.; Breiteneder, H., The latex-fruit syndrome. *Biochem Soc Trans* **2002,** *30* (Pt 6), 935-40.

29. (a) Allen, A. K.; Neuberger, A.; Sharon, N., The purification, composition and specificity of wheat-germ agglutinin. *Biochem J* **1973,** *131* (1), 155-62; (b) Sharon, N.; Lis, H., Lectins: cell-agglutinating and sugar-specific proteins. *Science* **1972,** *177* (4053), 949-59.

30. (a) Wright, C. S., Structural comparison of the two distinct sugar binding sites in wheat germ agglutinin isolectin II. *J Mol Biol* **1984,** *178* (1), 91-104; (b) Wright, C. S., 2.2 A resolution structure analysis of two refined N-acetylneuraminyl-lactose--wheat germ agglutinin isolectin complexes. *J Mol Biol* **1990,** *215* (4), 635-51; (c) Wright, C. S., Crystal structure of a wheat germ agglutinin/glycophorin-sialoglycopeptide receptor complex. Structural basis for cooperative lectin-cell binding. *J Biol Chem* **1992,** *267* (20), 14345-52; (d) Wright, C. S.; Jaeger, J., Crystallographic refinement and structure analysis of the complex of wheat germ agglutinin with a bivalent sialoglycopeptide from glycophorin A. *J Mol Biol* **1993,** *232* (2), 620-38.

31. Goldstein, I. J.; Poretz, R. D., The Lectins: Properties, Functions and Applications in Biology and Medicine. *Academic Press: San Diego, CA* **1986**, 33-247.

32. Burger, M. M.; Goldberg, A. R., Identification of a tumor-specific determinant on neoplastic cell surfaces. *Proc Natl Acad Sci U S A* **1967,** *57* (2), 359-66.

33. (a) Kronis, K. A.; Carver, J. P., Specificity of isolectins of wheat germ agglutinin for sialyloligosaccharides: a 360-MHz proton nuclear magnetic resonance binding study. *Biochemistry* **1982,** *21* (13), 3050-7; (b) Kronis, K. A.; Carver, J. P., Wheat germ agglutinin dimers bind sialyloligosaccharides at four sites in solution: proton nuclear magnetic resonance temperature studies at 360 MHz. *Biochemistry* **1985,** *24* (4), 826-33; (c) Wang, W. C.; Cummings, R. D., The immobilized leukoagglutinin from the seeds of Maackia amurensis binds with high affinity to complex-type Asn-linked oligosaccharides containing terminal sialic acid-linked alpha-2,3 to penultimate galactose residues. *J Biol Chem* **1988,** *263* (10), 4576-85.

34. Olsnes, S.; Kozlov, J. V., Ricin. *Toxicon* **2001,** *39* (11), 1723-8.

35. (a) van Damme, E. M. J.; Hao, Q.; Chen, Y.; Barre, F.; Vanderbussche, F.; Desmyter, S.; Rougé, P.; Peumans, J., Ribosome inactivating proteins: a family of plant proteins that do more than inactivate ribosomes. *Crit. Rev. Plant Sci.* **2001,** *20* (395); (b) Hung, C. H.; Lee, M. C.; Chen, J. K.; Lin, J. Y., Cloning and expression of three abrin A-chains and their mutants derived by site-specific mutagenesis in Escherichia coli. *Eur J Biochem* **1994,** *219* (1-2), 83-7; (c) Chen, J. K.; Hung, C. H.; Liaw, Y. C.; Lin, J. Y., Identification of amino acid residues of abrin-a A chain is essential for catalysis and reassociation with abrin-a B chain by site-directed mutagenesis. *Protein Eng* **1997,** *10* (7), 827-33.
36. (a) Gabius, H. J.; Darro, F.; Remmelink, M.; Andre, S.; Kopitz, J.; Danguy, A.; Gabius, S.; Salmon, I.; Kiss, R., Evidence for stimulation of tumor proliferation in cell lines and histotypic cultures by clinically relevant low doses of the galactoside-binding mistletoe lectin, a component of proprietary extracts. *Cancer Invest* **2001,** *19* (2), 114-26; (b) Andre, S.; Pieters, R. J.; Vrasidas, I.; Kaltner, H.; Kuwabara, I.; Liu, F. T.; Liskamp, R. M.; Gabius, H. J., Wedgelike glycodendrimers as inhibitors of binding of mammalian galectins to glycoproteins, lactose maxiclusters, and cell surface glycoconjugates. *Chembiochem* **2001,** *2* (11), 822-30.
37. Lutscha, G.; Nolla, F.; Ziskab, P.; Kindtb, A.; Franzb, H., Electron microscopic investigations on the structure of lectin I from Viscum album L. *FEBS Lett* **1984,** *21* (2), 335-338.
38. Niwa, H.; Tonevitsky, A. G.; Agapov, II; Saward, S.; Pfuller, U.; Palmer, R. A., Crystal structure at 3 A of mistletoe lectin I, a dimeric type-II ribosome-inactivating protein, complexed with galactose. *Eur J Biochem* **2003,** *270* (13), 2739-49.
39. Kaptein, R.; Dijkstra, K.; Nicolay, K., Laser photo-CIDNP as a surface probe for proteins in solution. *Nature* **1978,** *274* (5668), 293-4.
40. (a) Lee, R. T.; Gabius, H. J.; Lee, Y. C., Ligand binding characteristics of the major mistletoe lectin. *J Biol Chem* **1992,** *267* (33), 23722-7; (b) Lee, R. T.; Gabius, H. J.; Lee, Y. C., The sugar-combining area of the galactose-specific toxic lectin of mistletoe extends beyond the terminal sugar residue: comparison with a homologous toxic lectin, ricin. *Carbohydr Res* **1994,** *254*, 269-76; (c) Debray, H.; Montreuil, J.; Franz, H., Fine sugar specificity of the mistletoe (Viscum album) lectin I. *Glycoconj J* **1994,** *11* (6), 550-7; (d) Wu, A. M.; Chin, L. K.; Franz, H.; Pfuller, U.; Herp, A., Carbohydrate specificity of the receptor sites of mistletoe toxic lectin-I. *Biochim Biophys Acta* **1992,** *1117* (2), 232-4.
41. Jimenez-Barbero, J.; Dragoni, E.; Venturi, C.; Nannucci, F.; Arda, A.; Fontanella, M.; Andre, S.; Canada, F. J.; Gabius, H. J.; Nativi, C., Alpha-O-linked glycopeptide mimetics: synthesis, conformation analysis, and interactions with viscumin, a galactoside-binding model lectin. *Chemistry* **2009,** *15* (40), 10423-31.
42. Jiménez, M., PhD thesis. *Complutense University of Madrid (Madrid)* **2004**.
43. Ribeiro, J. P.; Andre, S.; Canada, F. J.; Gabius, H. J.; Butera, A. P.; Alves, R. J.; Jimenez-Barbero, J., Lectin-based drug design: combined strategy to identify lead compounds using STD NMR spectroscopy, solid-phase assays and cell binding for a plant toxin model. *ChemMedChem* **2010,** *5* (3), 415-9, 314.
44. Wu, A. M.; Song, S. C.; Wu, J. H.; Pfuller, U.; Chow, L. P.; Lin, J. Y., A sheep hydatid cyst glycoprotein as receptors for three toxic lectins, as well as Abrus precatorius and Ricinus communis agglutinins. *Biochim Biophys Acta* **1995,** *1243* (1), 124-8.
45. Muthing, J.; Meisen, I.; Bulau, P.; Langer, M.; Witthohn, K.; Lentzen, H.; Neumann, U.; Peter-Katalinic, J., Mistletoe lectin I is a sialic acid-specific lectin with

strict preference to gangliosides and glycoproteins with terminal Neu5Ac alpha 2-6Gal beta 1-4GlcNAc residues. *Biochemistry* **2004,** *43* (11), 2996-3007.
46. (a) Beuth, J.; Ko, H. L.; Gabius, H. J.; Burrichter, H.; Oette, K.; Pulverer, G., Behavior of lymphocyte subsets and expression of activation markers in response to immunotherapy with galactoside-specific lectin from mistletoe in breast cancer patients. *Clin Investig* **1992,** *70* (8), 658-61; (b) Hajto, T.; Hostanska, K.; Gabius, H. J., Modulatory potency of the beta-galactoside-specific lectin from mistletoe extract (Iscador) on the host defense system in vivo in rabbits and patients. *Cancer Res* **1989,** *49* (17), 4803-8.
47. Moisenovich, M.; Agapov, I.; Marx, U.; Bereiter-Hahn, J.; Tonevitsky, A., Intracellular transport of plant toxins ricin and viscumin from different plasma membrane sites. *Arzneimittelforschung* **2003,** *53* (6), 470-5.
48. Tonevitsky, A.; Agapov, I.; Chelnokova, O.; Moisenovich, M.; Marx, U., Comparison between the mechanisms of action of plant toxins ricin and viscumin on the stage of intracellular dissociation. *Arzneimittelforschung* **2002,** *52* (6), 500-5.
49. (a) Lambert, J. M.; Goldmacher, V. S.; Collinson, A. R.; Nadler, L. M.; Blattler, W. A., An immunotoxin prepared with blocked ricin: a natural plant toxin adapted for therapeutic use. *Cancer Res* **1991,** *51* (23 Pt 1), 6236-42; (b) Kreitman, R. J., Immunotoxins in cancer therapy. *Curr Opin Immunol* **1999,** *11* (5), 570-8; (c) Schmidt, A.; Mockel, B.; Eck, J.; Langer, M.; Gauert, M.; Zinke, H., Cytotoxic activity of recombinant bFGF-rViscumin fusion proteins. *Biochem Biophys Res Commun* **2000,** *277* (2), 499-506.
50. Kawaguchi, T.; Matsumoto, I.; Osawa, T., Studies on hemagglutinins from Maackia amurensis seeds. *J Biol Chem* **1974,** *249* (9), 2786-92.
51. Geisler, C.; Jarvis, D. L., Effective glycoanalysis with Maackia amurensis lectins requires a clear understanding of their binding specificities. *Glycobiology* **2011,** *21* (8), 988-93.
52. Arda, A.; Blasco, P.; Varon Silva, D.; Schubert, V.; Andre, S.; Bruix, M.; Canada, F. J.; Gabius, H. J.; Unverzagt, C.; Jimenez-Barbero, J., Molecular recognition of complex-type biantennary N-glycans by protein receptors: a three-dimensional view on epitope selection by NMR. *J Am Chem Soc* **2013,** *135* (7), 2667-75.
53. Abbas, A. K.; Lichtman, A. H., Cellular and Molecular Immunology. *Elsevier Saunders, Philadelphia, USA* **2005**.
54. Scrimgeour, M., Biochemistry. **1994**.
55. Mann, J.; Davidson, R. S.; Hobbs, J. B.; Banthorpe, D. V.; Harborne, J. B., In natural products: Their chemistry and biological significance. *Longman Group* **1994,** *1*.
56. (a) Perrin, C. L.; Armstrong, K. B.; Fabian, M. A., The origin of the anomeric effect: conformational analysis of 2-methoxy-1,3-dimethylhexahydropyrimidine. *J Am Chem Soc* **1994,** *116*, 715-722; (b) Lemieux, R. U.; Kullnig, R. K.; Bernstein, H. J.; Schneider, W. G., Configurational Effects on the Proton Magnetic Resonance Spectra of Six-membered Ring Compounds. *J Am Chem Soc* **1958,** *80* (22), 6098-6105.
57. Edward, J. T., Stability of Glycosides to Acid Hydrolysis. *Chem. Ind.* **1955**, 1102-1104.
58. Juaristi, E.; Cuevas, G., Recent Studies of the Anomeric Effect. *Tetrahedron* **1992,** *48*, 5019-5087.
59. Cocinero, E. J.; Carcabal, P.; Vaden, T. D.; Simons, J. P.; Davis, B. G., Sensing the anomeric effect in a solvent-free environment. *Nature* **2011,** *469* (7328), 76-9.
60. Vidal, P., Estudios sobre la conformaciíon de quitooligosacáridos y glycomiméticos del antígeno de Lewis A y su interacción con dominios de heveína y

galectinas. Una visión 3D utilizando RMN. *Estudios sobre la conformaciíon de quitooligosacáridos y glycomiméticos del antígeno de Lewis A y su interacción con dominios de heveína y galectinas. Una visión 3D utilizando RMN. PhD Thesis, Universidad Autónoma de Madrid* **2011**.

61. Claridge, T. D. W., High-Resolution NMR Techniques in Organic Chemistry. *Tetrahedron Organic Chemistry Series* **1999,** *19*.

62. (a) G.M., B.; Levy, H. A., Alpha-D-Glucose: Precise Determination of Crystal and Molecular

Structure by Neutron-Diffraction Analysis. *science* **1965,** *147*, 1038-1039; (b) Abeygunawardana, C.; Bush, C. A., Determination of the chemical structure of complex polysaccharides by heteronuclear NMR spectroscopy. *Adv Biophys Chem* **1993,** *3*, 199-249.

63. (a) von der Lieth, C.; Siebert, H.; Kozar, T.; Burchert, M.; Frank, M.; Gilleron, M.; Kaltner, H.; Kayser, G.; Tajkhorshid, E.; Bovin, N. V.; Vliegenthart, J. F.; Gabius, H., Lectin ligands: new insights into their conformations and their dynamic behavior and the discovery of conformer selection by lectins. *Acta Anat (Basel)* **1998,** *161* (1-4), 91-109; (b) Bush, C. A.; Martin-Pastor, M.; Imberty, A., Structure and conformation of complex carbohydrates of glycoproteins, glycolipids, and bacterial polysaccharides. *Annu Rev Biophys Biomol Struct* **1999,** *28*, 269-93; (c) Canales, A.; Angulo, J.; Ojeda, R.; Bruix, M.; Fayos, R.; Lozano, R.; Gimenez-Gallego, G.; Martin-Lomas, M.; Nieto, P. M.; Jimenez-Barbero, J., Conformational flexibility of a synthetic glycosylaminoglycan bound to a fibroblast growth factor. FGF-1 recognizes both the (1)C(4) and (2)S(O) conformations of a bioactive heparin-like hexasaccharide. *J Am Chem Soc* **2005,** *127* (16), 5778-9; (d) Casu, B.; Petitou, M.; Provasoli, M.; Sinay, P., Conformational flexibility: a new concept for explaining binding and biological properties of iduronic acid-containing glycosaminoglycans. *Trends Biochem Sci* **1988,** *13* (6), 221-5; (e) Marszalek, P. E.; Oberhauser, A. F.; Pang, Y. P.; Fernandez, J. M., Polysaccharide elasticity governed by chair-boat transitions of the glucopyranose ring. *Nature* **1998,** *396* (6712), 661-4.

64. (a) Mann, P. L.; Waterman, R. E., Glycocoding as an information management system in embryonic development. *Acta Anat (Basel)* **1998,** *161* (1-4), 153-61; (b) Noorman, F.; Barrett-Bergshoeff, M. M.; Rijken, D. C., Role of carbohydrate and protein in the binding of tissue-type plasminogen activator to the human mannose receptor. *Eur J Biochem* **1998,** *251* (1-2), 107-13; (c) Solis, D.; Estremera, D.; Usobiaga, P.; Diaz-Maurino, T., Differential binding of mannose-specific lectins to the carbohydrate chains of fibrinogen domains D and E. *Eur J Biochem* **1987,** *165* (1), 131-8.

65. Jimenez-Barbero, J.; Peters, T., NMR spectroscopy of glycoconjugates. *Wiley-VCH* **2003**.

66. Blundell, C. D.; Reed, M. A.; Overduin, M.; Almond, A., NMR spectra of oligosaccharides at ultra-high field (900 MHz) have better resolution than expected due to favourable molecular tumbling. *Carbohydr Res* **2006,** *341* (12), 1985-91.

67. Friebolin, H., Basic one- and two-Dimensional NMR spectroscopy. *Wiley-VCH* **2004**.

68. Becker, O. M.; Jr, A. D. M.; Roux, B.; Watanabe, M., Computational Biochemistry and Biophysics *Dekker Inc.* **2001**.

69. Morris, G. M.; Lim-Wilbi, M., Methods in Molecular Biology. **2008**, 365.

70. (a) Perez, S.; Imberty, A.; Engelsen, S. B.; Gruza, K.; Mazeau, K.; Jimenez-Barbero, J.; Poveda, A.; Espinosa, J. F.; Van Eyck, B. P.; Johnson, G.; French, A. D.;

Louise, M.; Kouwijzer, C. E.; Grootenius, P. D. J.; Bernardi, A.; Raimondi, L.; Senderowitz, H.; Durier, V.; Vergoten, G.; Rasmussen, K., A comparison and chemometric analysis of several molecular mechanics force fields and parameter sets applied to carbohydrates. *Carbohydr Res* **1998,** *314*; (b) Varma, C. K., Molecular Mechanical Force Fields. *Biochemistry* **2001,** *218*.
71. Pereira, C. S.; Kony, D.; Baron, R.; Muller, M.; van Gunsteren, W. F.; Hunenberger, P. H., Conformational and dynamical properties of disaccharides in water: a molecular dynamics study. *Biophys J* **2006,** *90* (12), 4337-44.
72. Case, D. A.; Darden, T. A.; Cheatham, I. T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B. P.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A., Universitiy of California, San Francisco. *Edition edn.* **2009**.
73. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A second generation force field for the simulation of proteins and nucleic acids. *J Am Chem Soc* **1995,** *117*, 5179-5197.
74. Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M., Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J* **2007,** *92* (11), 3817-29.
75. (a) Tessier, M. B.; Demarco, M. L.; Yongye, A. B.; Woods, R. J., Extension of the GLYCAM06 Biomolecular Force Field to Lipids, Lipid Bilayers and Glycolipids. *Mol Simul* **2008,** *34* (4), 349-363; (b) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem* **2008,** *29* (4), 622-55.
76. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J Comput Chem* **2004,** *25* (9), 1157-74.
77. Case, D. A.; Darden, T. A.; Cheatham, I. T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B. P.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A., AMBER 10. *University of California, San Francisco* **2008**.
78. (a) Hahn, E. L.; Maxwell, D. E., Spin Echo measurements of Nuclear Spin Coupling in Molecules. *Phys Rev* **1952,** *8* (5), 1070-1084; (b) McConnell, H. M., Reaction rates by Nuclear Magnetic Resonance. *J Chem Phys* **1958,** *2* (3), 430-431.
79. Meyer, B.; Peters, T., NMR spectroscopy Techniques for screening and Identifying Ligand Binding to Protein Receptors. *Angew Chem Int Ed Engl* **2003,** *42*, 864-890.
80. Nilsson, M.; Connell, M. A.; Davis, A. L.; Morris, G. A., Biexponential Fitting of Diffusion-Ordered NMR Data: Practicalities and Limitations. *Anal Chem* **2006,** *78* (9), 3040-3045.
81. (a) Montero, E.; Vallmitjana, M.; Pérez-Pons, J. A.; Querol, E.; Jiménez-Barbero, J.; Cañada, F. J., NMR studies of the conformation of thiocellobiose bound to

a β-glucosidase from Streptomyces sp. *FEBS lett* **1998,** *421* (3), 243-248; (b) Johnson, M. A.; Rotondo, A.; Pinto, B. M., NMR Studies of the Antibody-Bound Conformation of a Carbohydrate-Mimetic Peptide. *Biochemistry* **2002,** *41* (7), 2149-2157; (c) Roldós, V.; Cañada, F. J.; Jimenez-Barbero, J., Carbohydrate-Protein Interactions; A 3D view by NMR. *Chembiochem* **2011,** *12*, 1-17.

82. Akasaka, K., Intermolecular spin diffusion as a method for studying macromolecule-ligand interactions. *J Magn Reson* **1979,** *36* (1), 135-140.

83. (a) Bernardi, A.; Potenza, D.; Capelli, A. M.; Garcia-Herrero, A.; Canada, F. J.; Jimenez-Barbero, J., Second-generation mimics of ganglioside GM1 oligosaccharide: a three-dimensional view of their interactions with bacterial enterotoxins by NMR and computational methods. *Chemistry* **2002,** *8* (20), 4597-612; (b) Bernardi, A.; Arosio, D.; Potenza, D.; Sanchez-Medina, I.; Mari, S.; Canada, F. J.; Jimenez-Barbero, J., Intramolecular carbohydrate-aromatic interactions and intermolecular van der Waals interactions enhance the molecular recognition ability of GM1 glycomimetics for cholera toxin. *Chemistry* **2004,** *10* (18), 4395.

84. (a) Jayalakshmi, V.; Rama Krishna, N., Complete Relaxation and Conformational Exchange Matrix (CORCEMA) Analysis of Intermolecular Saturation Transfer Effects in Reversibly Forming Ligand–Receptor Complexes. *J Magn Reson* **2002,** *155* (1), 106-118; (b) Angulo, J.; Enriquez-Navas, P. M.; Nieto, P. M., Ligand-receptor binding affinities from saturation transfer difference (STD) NMR spectroscopy: the binding isotherm of STD initial growth rates. *Chemistry* **2010,** *16* (26), 7803-12.

85. Tsvetkov, Y. E.; Burg-Roderfeld, M.; Loers, G.; Arda, A.; Sukhova, E. V.; Khatuntseva, E. A.; Grachev, A. A.; Chizhov, A. O.; Siebert, H. C.; Schachner, M.; Jimenez-Barbero, J.; Nifantiev, N. E., Synthesis and molecular recognition studies of the HNK-1 trisaccharide and related oligosaccharides. The specificity of monoclonal anti-HNK-1 antibodies as assessed by surface plasmon resonance and STD NMR. *J Am Chem Soc* **2012,** *134* (1), 426-35.

86. (a) Kover, K. E.; Groves, P.; Jimenez-Barbero, J.; Batta, G., Molecular recognition and screening using a 15N group selective STD NMR method. *J Am Chem Soc* **2007,** *129* (37), 11579-82; (b) Feher, K.; Groves, P.; Batta, G.; Jimenez-Barbero, J.; Muhle-Goll, C.; Kover, K. E., Competition saturation transfer difference experiments improved with isotope editing and filtering schemes in NMR-based screening. *J Am Chem Soc* **2008,** *130* (50), 17148-53; (c) Diercks, T.; Ribeiro, J. P.; Canada, F. J.; Andre, S.; Jimenez-Barbero, J.; Gabius, H. J., Fluorinated carbohydrates as lectin ligands: versatile sensors in 19F-detected saturation transfer difference NMR spectroscopy. *Chemistry* **2009,** *15* (23), 5666-8; (d) André, S.; Cañada, F. J.; Shiao, T. C.; Largartera, L.; Diercks, T.; Bergeron-Brlek, M.; el Biari, K.; Papadopoulos, A.; Ribeiro, J. P.; Touaibia, M.; Solís, D.; Menéndez, M.; Jiménez-Barbero, J.; Roy, R.; Gabius, H.-J., Fluorinated Carbohydrates as Lectin Ligands: Biorelevant Sensors with Capacity to Monitor Anomer Affinity in 19F-NMR-Based Inhibitor Screening. *Eur J Org Chem* **2012,** *4*, 4354-4364.

Chapter 2. Molecular Recognition of oligosaccharides related to the SF3a O-Antigen by IgG and IgM monoclonal antibodies

2.1 Introduction

Shigellosis or bacillary dysentery is an invasive and highly contagious disease of the human colon. In fact, it is one among the many forms of enteric infections that to a large extent affects worldwide[1]. It causes 2 billion cases in humans per year[2], being severe and even fatal in children between one and five years.

Shigella flexneri is the most frequently isolated Shigellosis causative agent, especially in developing countries. The protective host immune response to *S. flexneri* is against the O-Antigen part of the lipopolysaccharide of the outer membrane of the bacterium.

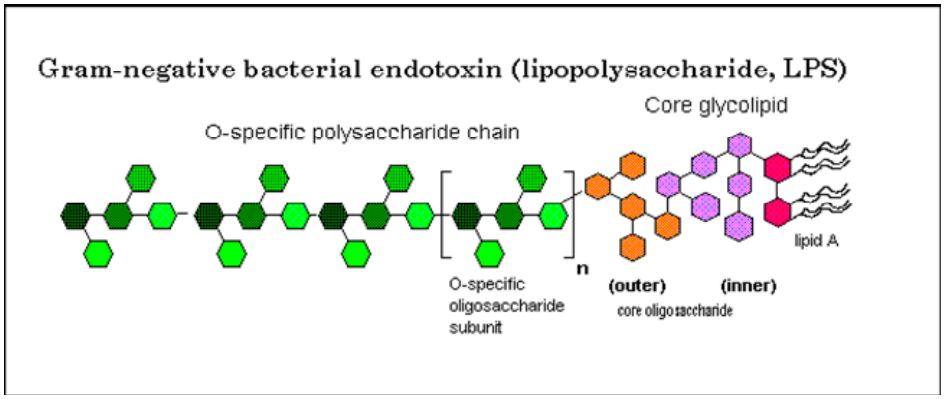


Figure 2.1 Schematic representation of the structure of the Lipopolysacchairde of a GRAM-negative bacterial endotoxin.

The lipopolysaccharide (LPS, See Figure 2.1), which is produced by Gram-negative bacteria, is a powerful activator of innate immune responses. They are the major components of the outer membrane of GRAM-Negative bacteria and exert, in isolated form, a variety of biological activities in animals and plants. LPS is a powerful activator of innate immune responses and is responsible for endotoxic shock, which is the often fatal complication of sepsis. It is only in the past few years that the molecular mechanisms by which LPS initiates signaling responses in immune system have been elucidated [3].

The general structure of bacterial LPS consists of a hydrophobic lipid A domain, an oligosaccharide 'core' and a distal polysaccharide (or O-Antigen)[3-4].

The lipid A moiety alone is sufficient to activate the innate immune response; adaptive (antibody) responses are generated to the O-antigen polysaccharide later in the course of an infection. Lipid A consists of a diglucosamine diphosphate head group that is substituted with a variable number of acyl chains, ranging from four to eight[3].

The outer core region provides an attachment site for the O-polysaccharide (O-Antigen)[4]. In bacteria the core oligosaccharides are conceptually divided into two regions: inner core (lipid A proximal) and outer core. Bacteria produce lipooligosaccharides (LOSs) that contain a recognizable inner core from which extend one or more mono- oligosaccharide branches (equivalent to the outer core). Within a genus or family, the structure of the inner core tends to be well conserved, and the fact that the core oligosaccharides from distantly related bacteria share structural features in the inner core is a reflection of the importance of the core in outer-membrane integrity. The inner core typically contains residues of Kdo and L-glycerol-D-Mannoheptose (L,D-Hhep). The last Kdo linked is often modified with a phosphate or ethanolamine group. The outer core is made of hexose residues that are attached to the last heptose residue in the inner core[4]. Hexoses often found in the outer core include D-Glc, D-Man, D-Gal, etc. There is usually at least three hexoses linked β1→3, with the O-Antigen being attached to the third hexose.

A repetitive glycan polymer contained is referred to as the O-Antigen, O-Polysaccharide, or O-Side-chain of the bacteria. The location of O-polysaccharide at the cell surface places it at the interface between the bacterium and its environment. The O-Antigen is attached to the core oligosaccharide, and comprises the outermost domain of the LPS molecule. The composition of the O-chain varies from strain to strain. For example, there are over 160 different O-antigen structures produced by different *E. coli* strains[4]. The O-Antigen is exposed on the outer surface of the bacterial cell, and, as a consequence, is a target for recognition by host antibodies. The O-Antigen is composed by repeated units (RUs) that can differ in the monomer glycoses, the position and stereochemistry of the O-glycosidic linkages, and the presence or absence of non-carbohydrate substituents. O-RUs from different structures may comprise varying numbers of monosaccharides, they may be linear or branched. In some cases, non stoichiometric modifications (e.g.,O-Acetylation or glycosylation)[5] have been identified.

Early interest in the chemistry, biosynthesis, and genetics of O-Antigens was stimulated by their roles as essential virulence determinants and their potential application in vaccine development, but information is now available for a wide range of species with different lifestyles. The primary role(s) of the O-polysaccharides appears to be protective. In animal pathogens, O-polysaccharides may contribute with their

chemistry (chain length and sugars composition) to bacterial evasion of host immune responses, particularly the alternative complement cascade[6].

In the case of *Shigella flexneri*, depending on the structure of the O-Antigen, there are 14 different serotypes. Every serotype contains repeated units of the same tetrasaccharide backbone[2, 7] (See Fig. 2.2). The differences between the serotypes are conferred by phage modifications, including addition to the basic structure of α-D-glucose or/and O-Acetyl groups. These modifications are immune determinants within *S. flexneri* serotypes[7]. These substitutions result from the action of α-D-glucopyranosyl and/or O-acetyltransferases coded by serotype-converting bacteriophages inserted in the bacterium chromosome.

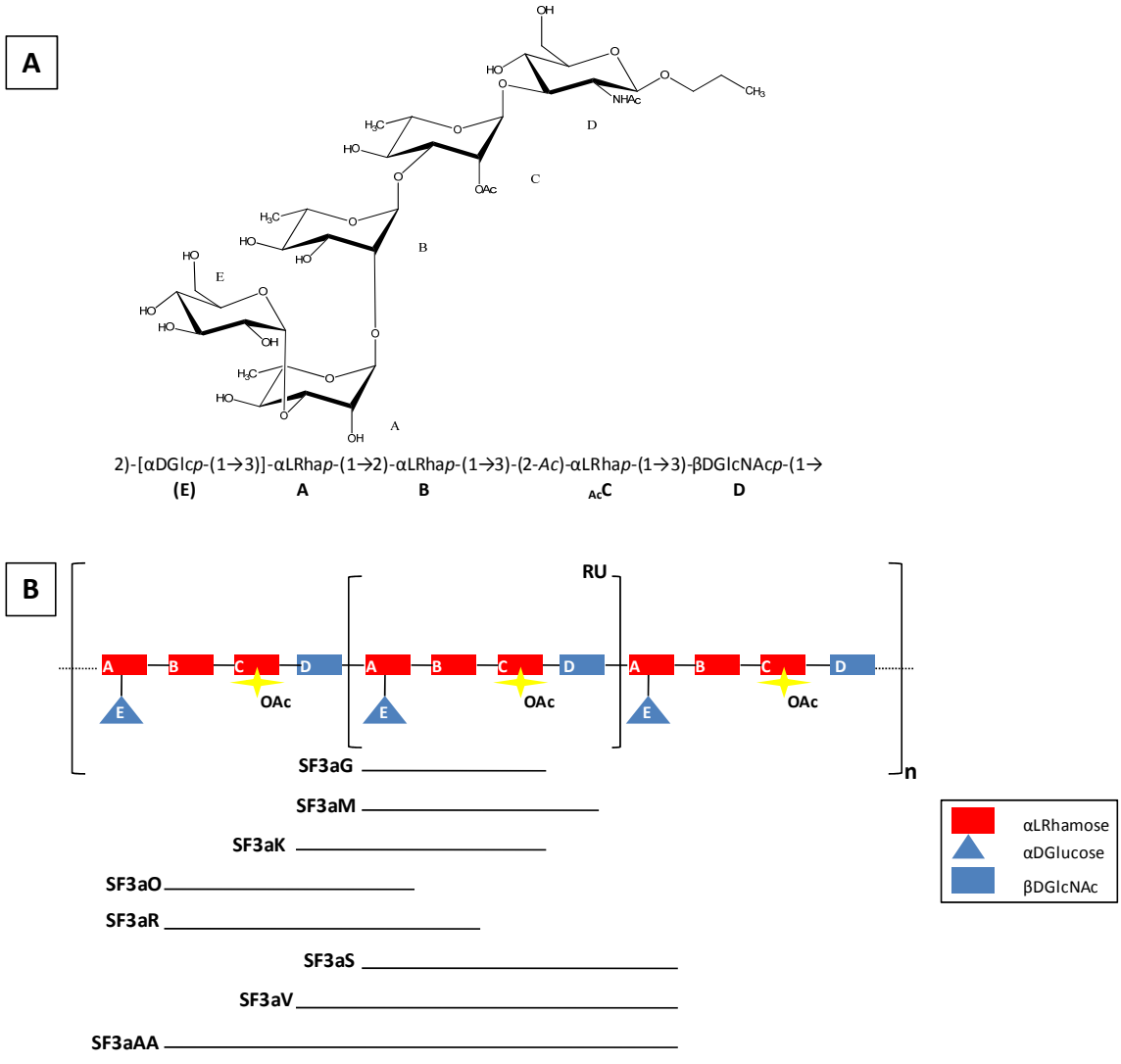


Figure 2.2 (A) SF3a synthetic oligosaccharide backbone used in the study. (B) Shematic representation of the oligosaccharides used in the study.

This serotype diversity and virulence supposes a top priority of WHO (World Health organization) implying the development of a potent vaccine that can protect

against all bacterium serotypes. From inspection of the chemical features of the different group types and after type factor characterization, it was assumed that immunizing with O-antigens from only serotypes 3a, 2a and 6 could protect against almost every *S. flexneri* serotypes[8]. Indeed, it can be envisaged that acquiring detailed knowledge of the fine chemical details of the three-dimensional shapes of these O-antigens is essential for a definitive substantiation of the sero-specificity of strains at the molecular level, as well as for a better understanding of the role of the O-antigens in the final vaccine formulation[7].

Understanding at the maximum possible resolution the interactions of these antigens with specific monoclonal antibodies (mAb) should be a key step for designing more effective O-antigen based vaccines[9]. Moreover, it could permit to exquisitely define the specificity of the isolated polysaccharides and the possibility of being replaced by well-defined synthetic carbohydrates.

On this basis, it has been shown that modifications of the O-antigen structure by either glucosylation or O-Acetylation enhance the survival of *S. flexneri*[5]. According to the existing data, it has been deduced that glucosylation modulates the polysaccharide conformation to acquire a more compact structure, then increasing its exposition at the host cell surface. On the other hand, the fine structural-activity details of the role of the O-Acetylation in pathogenesis caused by *S. flexneri* remains still unknown.

Nevertheless, it has been postulated that the exceptionally high serological impact [5, 10] of the 2-O-Acetyl group at Rhamnose C in SF3a serotype (See Fig. 2.2) may be accounted for by its axial orientation. In fact, a striking difference was observed when the immunochemical specificities of native O-Acetylated versus de-O-Acetylated SF3a LPS were tested against SF3a antisera[10]. None of them were bound to de-O-Acetylated LPS. This result permitted to propose that the 2-O-axially oriented presentation is essential for the recognition with receptors and antibodies as compared with those presented by O-Acetylation at the other positions.

Evidences on the molecular recognition features of other *S. flexneri* serotypes have been presented. For instance, for SF2a serotype,[9] the binding epitope of differently modified oligosaccharides (AB(E)CD, B(E)CDA and ECDA'B'), towards the protective monoclonal antibody F22-4 has been elucidated. It was shown that trisaccharide ECD (E, C and D are αDGlc*p*, αLRha*p* and βDGlcNAc*p* respectively) is

the key moiety that specifically interacts with the mAb. NMR experiments assisted by Molecular Dynamics simulations showed the existence of certain glycosidic conformational restriction at the ECD motif upon binding, although the terminal residues remained rather flexible. Additionally, mimics of the *S. flexneri* 5a O-specific polysaccharide [11] have been employed to improve their specificity towards anti O-antigen antibodies. For instance, the presence of a branched Glucose (residue E) taking part of the DA(E)BC pentasaccharide essential for its recognition by the protective mIgA C5 and I3 antibodies, while for *S. flexneri* type 3a, the key structure is composed by one α-D-Glc residue linked to a 3-O-Rha A residue, while O-2 of Rha C is O-acetylated (Fig. 2.2A).

2.2 Objectives

Herein, we have dissected the fine conformational and molecular details of the interaction between a series of oligosaccharides related to the original O-Antigen and specific monoclonal antibodies. In particular, a combination of experimental (STD-NMR[12] and trNOESY-NMR) and computational methods (MD simulations) has been employed to gain the key structural insights into their specificity. The main object of the project is to study the ability of synthetically available SF3a serotype oligosaccharides to mimic the original *O*-antigen and deduced their molecular recognition and conformational behavior upon binding to specific mAb IgG C7-37 and IgM G19-2 antibodies. The strategy may facilitate the design of either ligands or carbohydrate recognition domains important in the design of vaccines against *S. flexneri* serotype 3a (SF3a).

The synthetic oligosaccharides employed in this study are based on the natural RU structure of the SF3a serotype (EABCD, dubbed as SF3aM, see Fig 2.2A). The employed oligosaccharides (See Fig. 2.2B) display different substitutions of the same natural SF3a RU sequence (EABC, EABCD, D(E)ABC, BCDAE, B'CD(E)AB, E'A'BCD(E)A, D'(E')A'BCD(E)A, B'C'D'(E')A'BCD(E)A for the SF3aG, M, K, O, R, S, V, and AA synthetic oligosaccharides respectively). The synthetic oligosaccharides and specific antibodies IgG C7-37 and IgM G19-2 were provided by Dr. Laurence Mulard's group at Institute Pasteur (Paris).

2.3 The conformation of the oligosaccharides in the free state; *NMR and MD simulations*

The MD simulation data provided ensemble-averaged geometries that were employed to measure the key interprotonic interglycosidic distances that could give rise to NOE/ROE correlations. Then, they were compared to the NOESY/ROESY experimental data. On this side, the analysis of the cross peaks in these NOE-based spectra allowed establishing the experimental average interprotonic distances along the glycosidic linkages.

From now on, the behavior of the different linkages will be presented:

The E-A linkage.-This glycosidic linkage (See Fig. 2.2) is present in all ligands. According to the MD simulations, there are populations of two conformers, which differ in the orientation of Ψ angle (*syn*Ψ- and *syn*Ψ+), while Φ is always dictated by the exo-anomeric effect. Fittingly, according the MD simulations, their relative populations depended on the chemical substitution at the vicinity of the linkage. In particular, for the non-reducing end of ligands SF3aG, SF3aM and SF3aS, the MD simulations predicted a major *syn*Ψ- geometry (ca. 60% of the population) and a minor *syn*Ψ+ conformation. The key NOEs are H1E-H2A and H1E-H3A. Their intensities may be translated in the relative populations of both conformers. In contrast, for the reducing end of SF3aK, SF3aO, SF3aR SF3aV SF3aAA, and SF3aS, the presence of a D residue attached at O2 of Rhap A affected the conformational behavior around E-A glycosidic linkage. Now, the presence of the *syn*Ψ- conformation is further increased, while the alternative *syn*Ψ+ conformation showed the concomitant decrease in population. Inspection of molecular models suggested that this geometry is now more sterically-hindered[8]. The presence of residue D also provides one additional NOE constrain. Only when the conformation around the glycosidic torsion angles of the E-A fragment is exoΦ/*syn*Ψ- (and also at the D-A disaccharide moiety, see later), there is a short H1D-H1E distance. Therefore, this H1D-H1E is exclusive for the exoΦ/*syn*Ψ- geometry around the E-A glycosidic linkage (See Fig. 2.3).

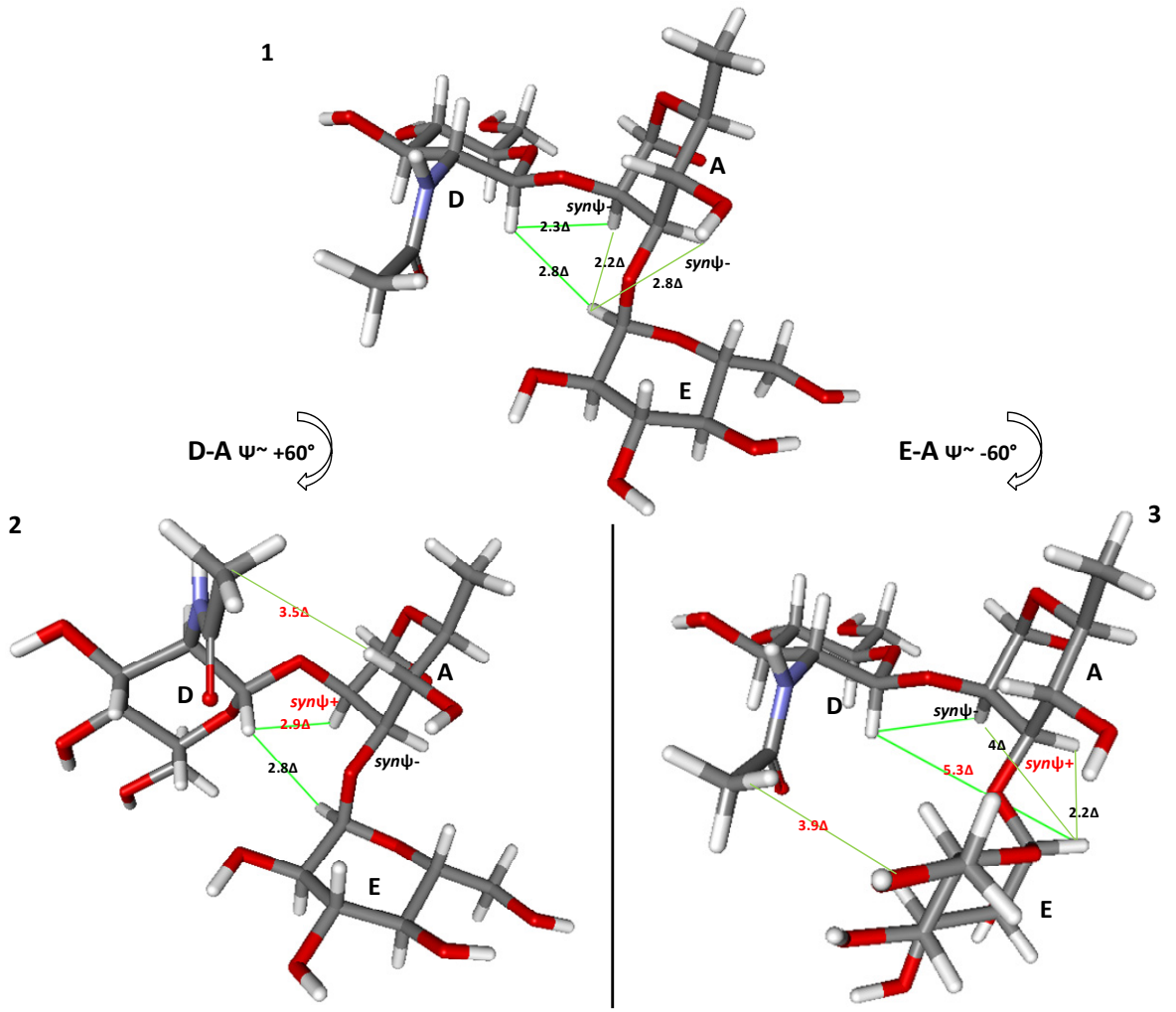


Figure 2.3 Detail of E-A-D fragment of SF3aK oligosaccharide showing the relevant distances for every possible conformation around E-A and D-A glycosidic torsions.

Indeed, the experimental NMR distances (Tables 2.1 A-H) derived from the ROESY spectra were, in cases, in very satisfactory agreement with the MD-based data. Thus, in the free state for all SF3a ligands, there are two conformational families in solution around the A-E linkage, exoΦ/*syn*Ψ+ and exoΦ/*syn*Ψ-, with a preference for the last one, which is further increased when residues D-like are present. The H1D-H1E NOE was always present for SF3aK, SF3aO, SF3aR, SF3aS, SF3aV, and SF3aAA. In any case, the presence of certain conformational mobility around E-A (and E'-A') torsions is granted from the combined MD/NMR approach. For example, in the case of ROESY spectrum of SF3aM oligosaccharide (See Figure 2.5), the strong interresidual H1E-H2A and H1E-H3A cross peaks observed are both stronger than the intraresidual H1E-H2E cross peak. Therefore, the presence of a conformational equilibrium is granted.

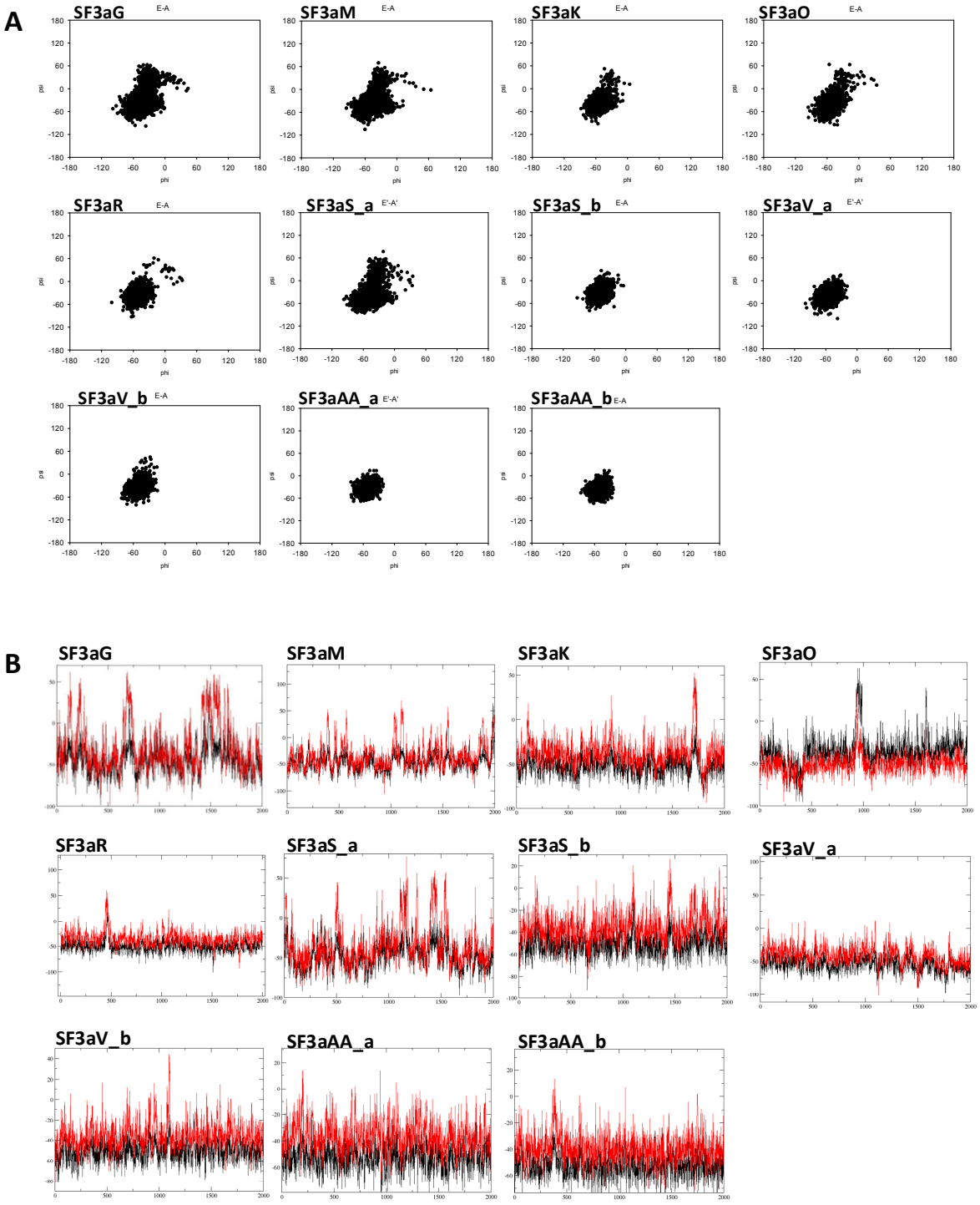


Figure 2.4 (A) Plots of the available conformational space for the Φ/Ψ glycosidic linkages corresponding to the E/A linkage. The presence of a major syn-Ψ- conformation may be expected, especially for the D-substituted molecules. (B) Φ/Ψ (showed in Black/Red lines respectively) values of E/A glycosidic torsion of each SF3a oligosaccharides during the MD simulation.

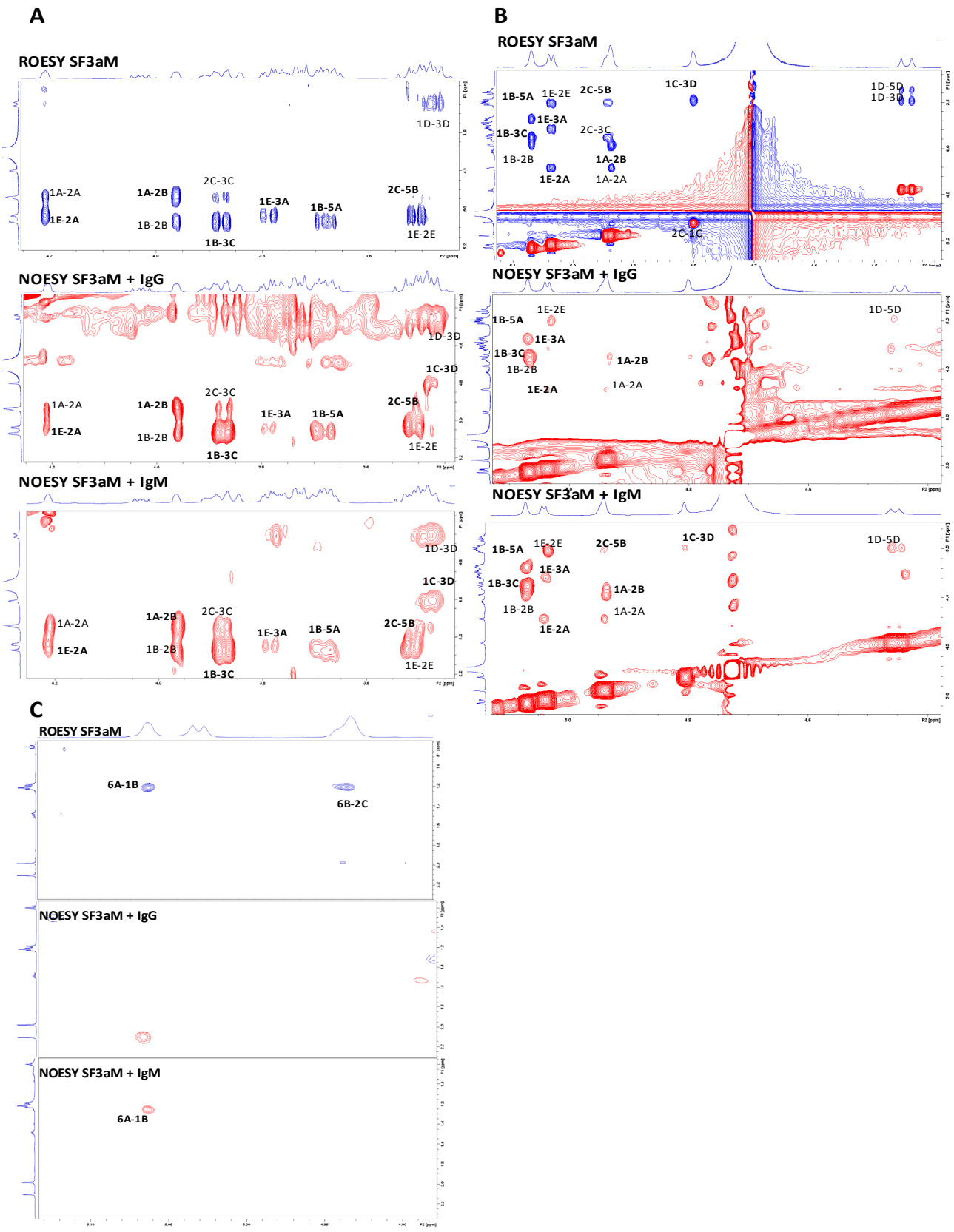


Figure 2.5 The key NOEs from relevant zones in the spectra found in the F2 (A, C) and F1 (B) dimension for SF3aM ligand. The ROESY spectra in the free state are shown always at the top (in blue) and the trNOESY in the presence of IgG (center) or IgM (below) in the trNOESY spectra are also shown.

The D-A linkage.-According to the MD data, this torsion angle could sample two conformational regions (exoΦ/*syn*Ψ+ and exoΦ/*syn*Ψ-), with a major population of the exoΦ/*syn*Ψ+ geometry. In the studied ligands, the D-A linkage is always associated with the presence of one E residue attached to the A moiety, which influences the conformational behavior. Indeed, there is a correlation between the E-A and D-A glycosidic torsions, since a simultaneous exoΦ/*syn*Ψ+ orientation for both E-A and D-A linkages is satirically hindered. In fact, the MD simulations for the different oligosaccharides showed that the D-A(E) fragment may adopt three different geometries. The exoΦ form is always present, while the two Ψ angles may be present both as *syn*Ψ- or a combination of *syn*Ψ- and *syn*Ψ+ for the D-A and E-A linkages, as outlined in Figure 2.3. However, as mentioned above, the presence of the H1D-H1E NOE strongly suggests that the exoΦ/*syn*Ψ- conformer around E-A is present. The actual presence of the different conformers was assessed from the comparison of the expected distances for the three conformations, estimated form MD simulations with those estimated from the NOE measurements (Table 2.1 A-H). While the H1D-H2A interglycosidic distance is related to the conformation around this particular glycosidic linkage (2.3-2.4Å for the *syn*Ψ- and longer, around 2.9Å for the *syn*Ψ+ forms), the presence or absence of the remote H1D-H1E and NHAcD-H3E/H4A long-range distances is related to the combination of particular values for both glycosidic torsions. For instance, for a double *syn*Ψ- conformation or *syn*Ψ- and *syn*Ψ+ around D-A and E-A, respectively, a value of 2.8 Å for H1D-H1E would be expected (with an observable NOE). The other *syn*Ψ+/*syn*Ψ- combination for D-A and A-E, respectively would provide much longer distances (Fig 2.3). The global comparison between the MD-based distance values and those deduced from the NMR data are gathered in Tables 2.1 A-H. For example, in the case of ROESY spectrum of SF3aO oligosaccharide (See Figure 2.7), the interresidual H1D-H2A and H1E-H1D cross peaks are observed, weaker than the intra-residual H1D-H3D one. Therefore, according to the obtained data, the NOE data strongly suggest the presence of a conformational equilibrium around Ψ of this D-A (or D'-A') glycosidic linkage.

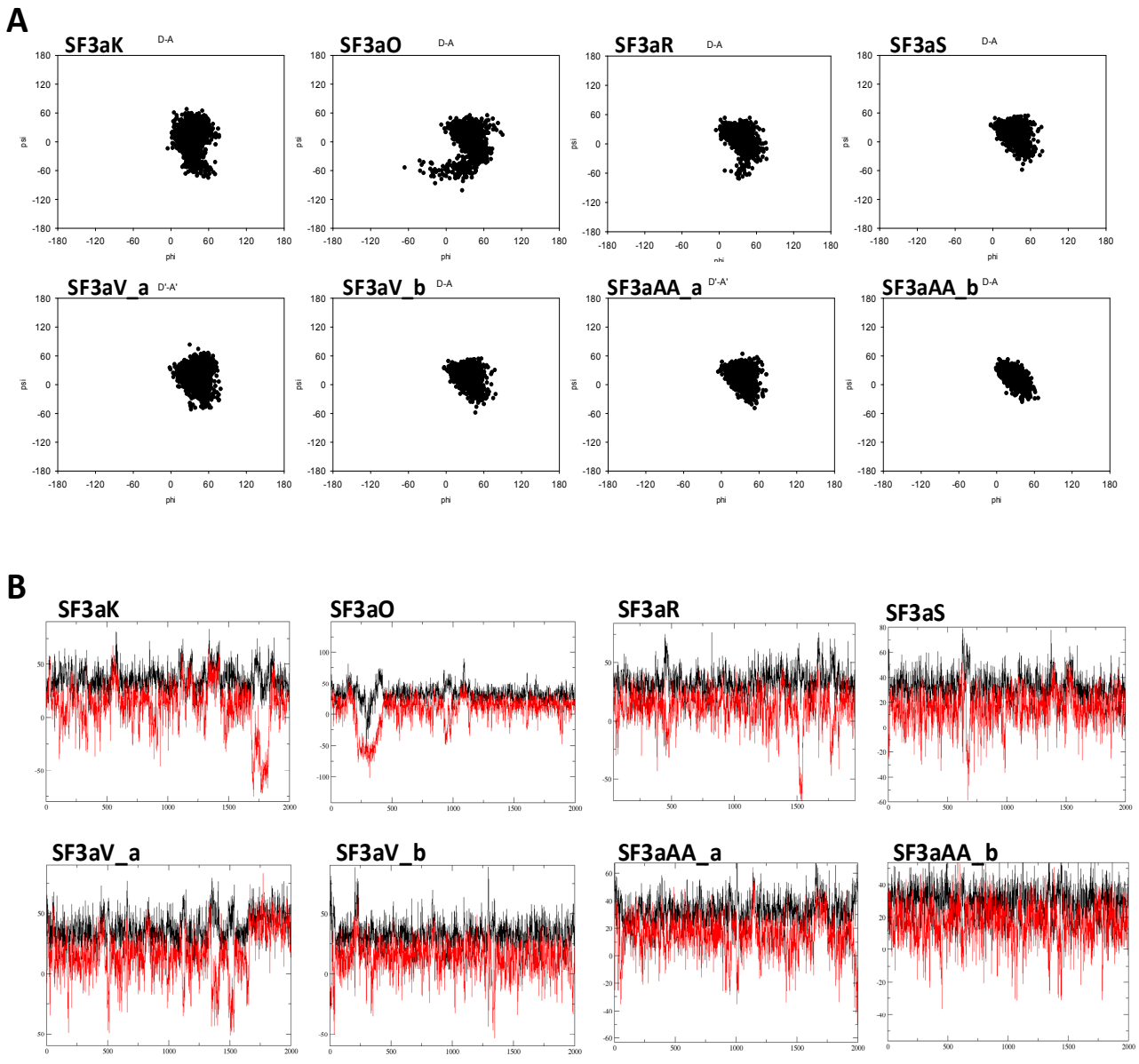


Figure 2.6 (A) Plots of the available conformational space for the Φ/Ψ glycosidic linkages corresponding to the D/A linkage. The presence of a conformational equilibrium may be expected. (B) Φ/Ψ (showed in Black/Red lines respectively) values of D/A glycosidic torsion of each SF3a oligosaccharides during the MD simulation.

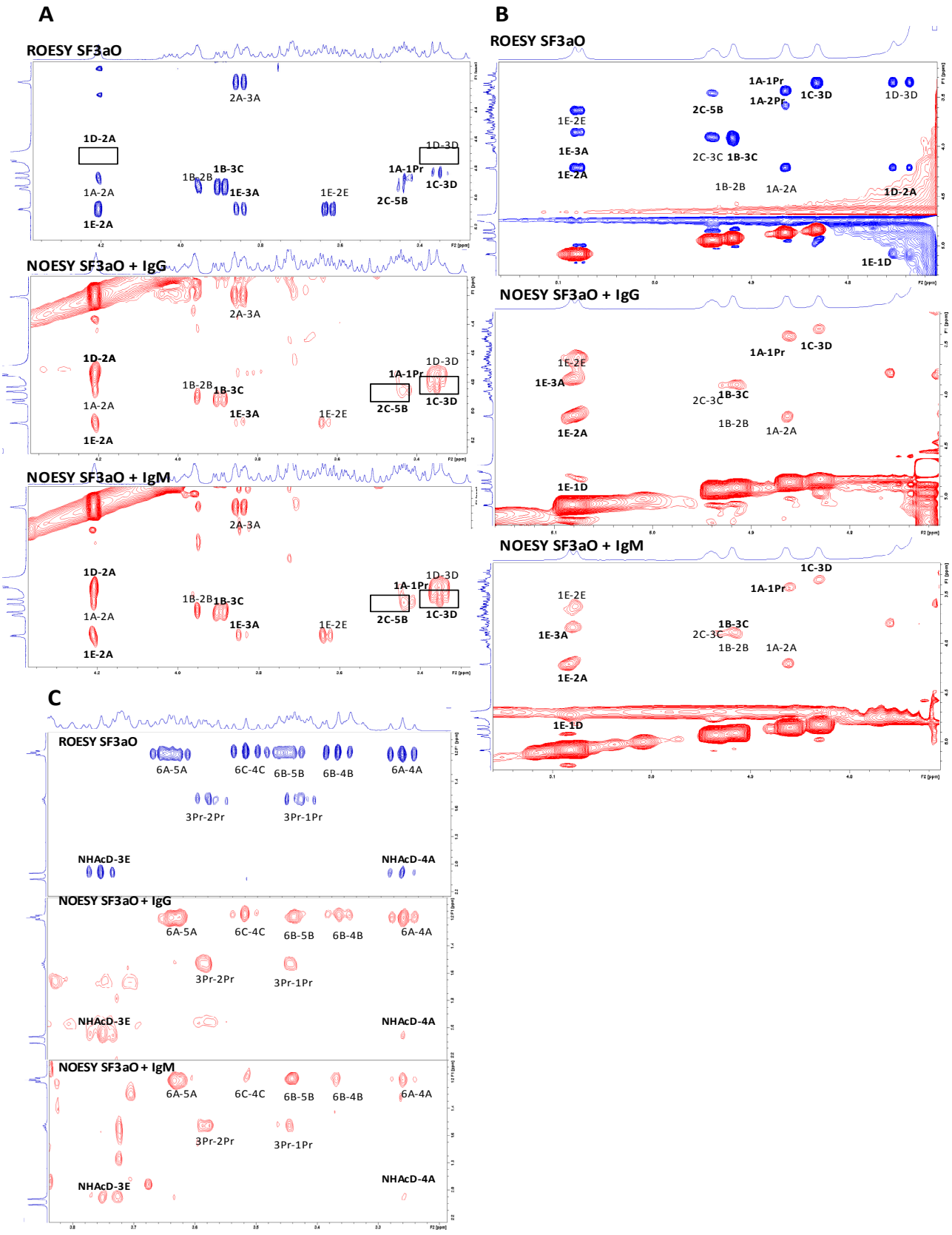


Figure 2.7 The key NOEs from relevant zones in the spectra found in the F2 (A, C) and F1 (B) dimension for SF3aO ligand. The ROESY spectra in the free state are shown always at the top (in blue) and the trNOESY in the presence of IgG (center) or IgM (below) in the trNOESY spectra are also shown.

The B-C linkage.-The B-C glycosidic linkage is present in all ligands. The MD simulations suggest that two conformations around the glycosidic linkage (exoΦ/*syn*Ψ+ and exoΦ/*syn*Ψ-) are present with similar populations. However, the distances derived from ROESY spectra strongly indicated that all these SF3a ligands show a strong preference for the exoΦ/*syn*Ψ- geometry (See Tables 2.1 A-H). For example, in the case of ROESY spectrum of SF3aS oligosaccharide (See Figure 2.9), the interresidual H1B-H3C cross peak is observed, stronger than the intraresidual H1B-H2B one. Therefore, the presence of a major exoΦ/*syn*Ψ- geometry can be deduced. In particular, the presence of a CH36B-H2C NOE is exclusive of this conformation (associated to a distance of ~3.3Å), together with the occurrence of a very strong H1B-H3C NOE (distance of 2.1-2.2Å) stronger than that between H5B and H2C (2.5 Å) supports this conclusion.

The C-D linkage.-The C-D glycosidic torsion is present in ligands SF3aM, SF3aO, SF3aR, SF3aS, SF3aV and SF3aAA. According to the MD simulations, the conformational behavior of C-D strongly depends on the particular environment. For non-substituted D GlcNAc residues, as in SF3aM, the conformational space around the C-D linkage is fairly wide, with similar populations of the two possible exoΦ/*syn*Ψ+ and exoΦ/*syn*Ψ- geometries. However, according to the MD, the presence of one Rhap A linked to D in the rest of the ligands affects its conformational behavior and the exoΦ/*syn*Ψ+ conformation is now somehow more populated[8]. These MD-based predictions were in close agreement with the analysis of the key inter-proton distances estimated from the ROESY experiments. For both exoΦ/*syn*Ψ+ and exoΦ/*syn*Ψ- geometries, it is expected a short H1C/H3D distance (2.3 Å). In principle, there is also an exclusive H5C/H4D NOE only for the exoΦ/*syn*Ψ- conformation (2.7 Å), but it was not found. For example, in the case of ROESY spectrum of SF3aR oligosaccharide (See Figure 2.11), the interresidual H1C-H3D cross peak is observed. The obtained data suggest the presence of a conformational equilibrium around Ψ.

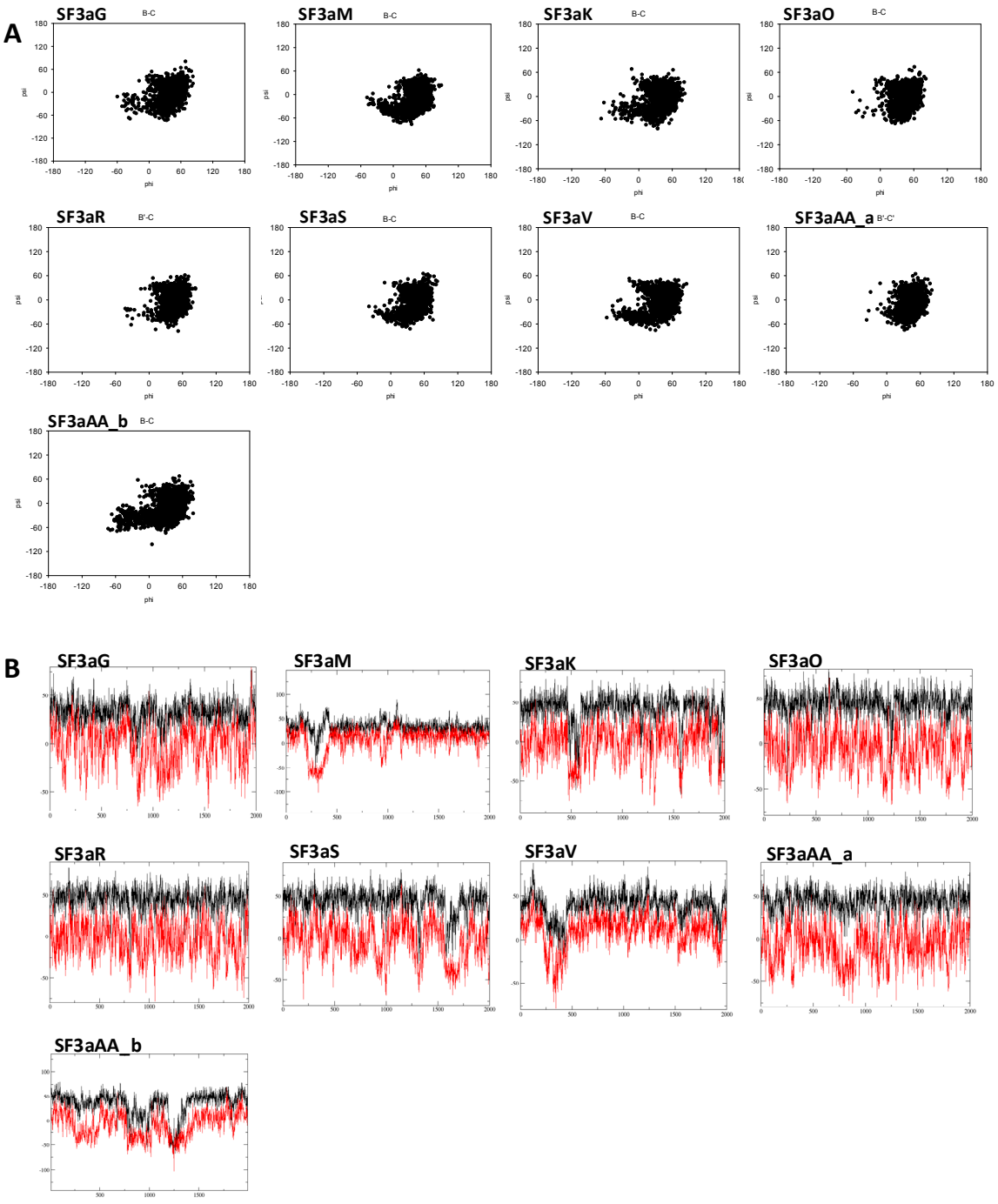


Figure 2.8 (A) Plots of the available conformational space for the Φ/Ψ glycosidic linkages corresponding to the B/C linkage. The presence of a conformational equilibrium may be expected. (B) Φ/Ψ (showed in Black/Red lines respectively) values of B/C glycosidic torsion of each SF3a oligosaccharides during the MD simulation.

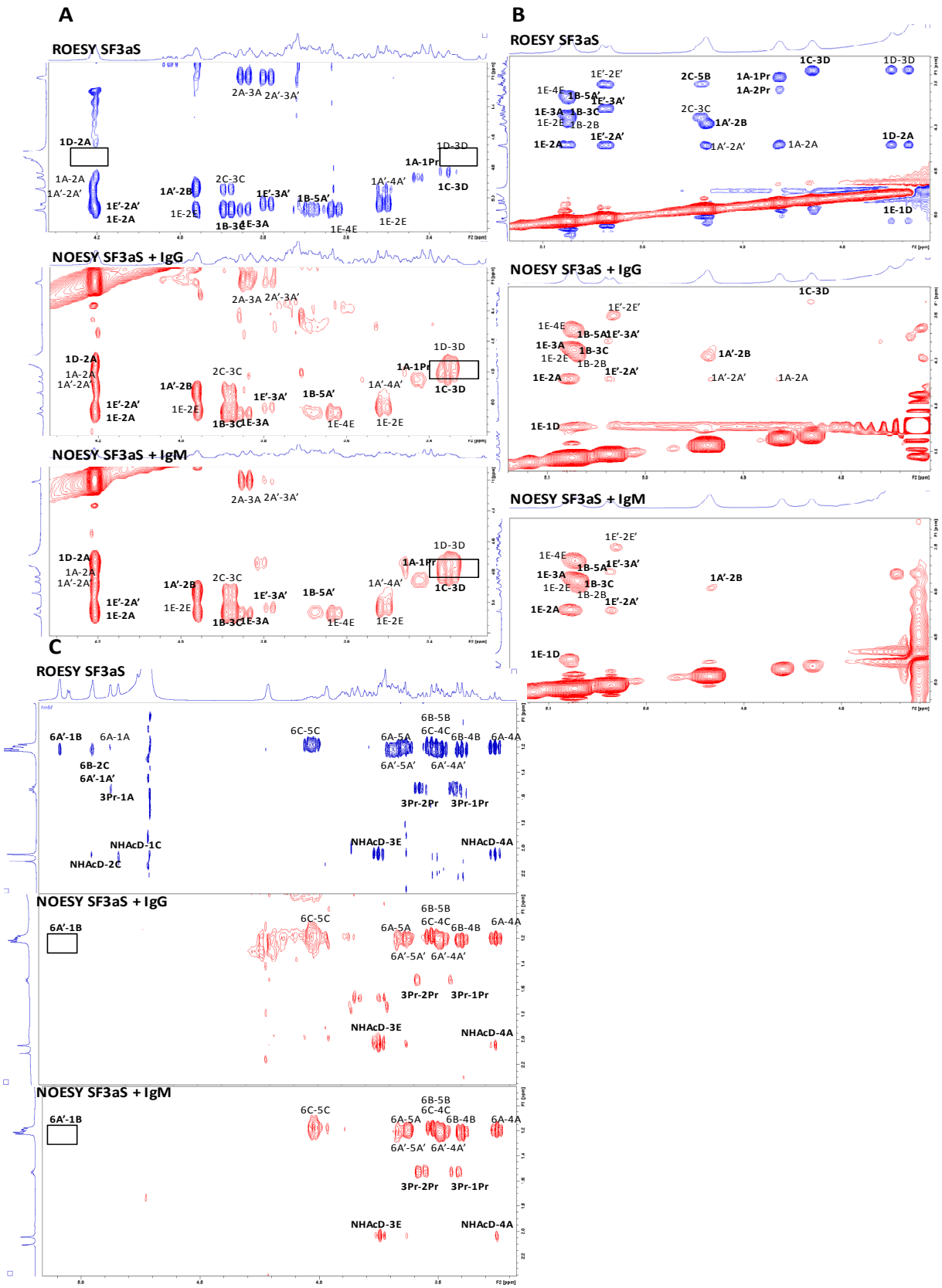


Figure 2.7 The key NOEs from relevant zones in the spectra found in the F2 (A, C) and F1 (B) dimension for SF3aS ligand. The ROESY spectra in the free state are shown always at the top (in blue) and the trNOESY in the presence of IgG (center) or IgM (below) in the trNOESY spectra are also shown.

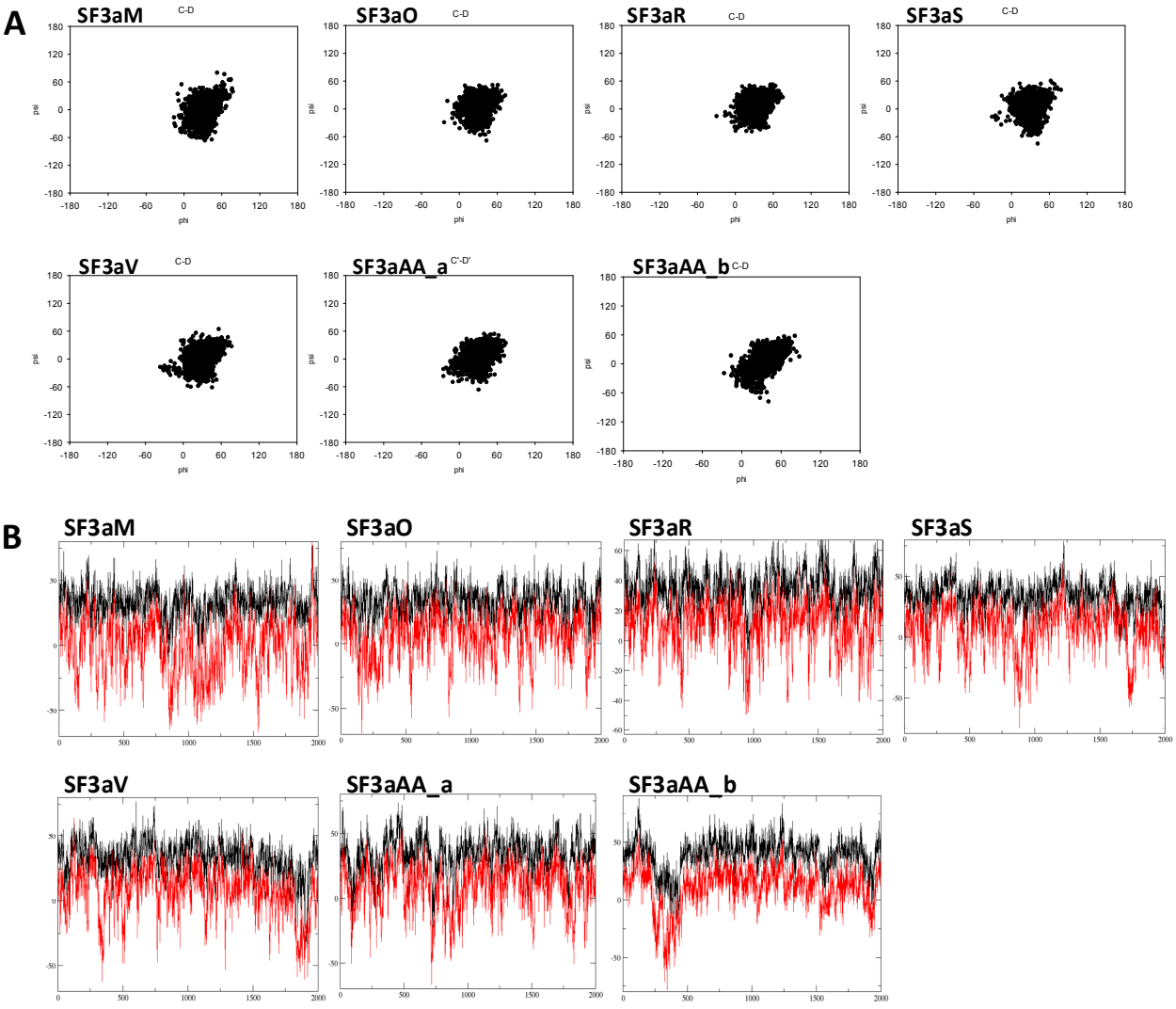


Figure 2.10 (A) Plots of the available conformational space for the Φ/Ψ glycosidic linkages corresponding to the C/D linkage. The presence of a conformational equilibrium may be expected. (B) Φ/Ψ (showed in Black/Red lines respectively) values of C/D glycosidic torsion of each SF3a oligosaccharides during the MD simulation.

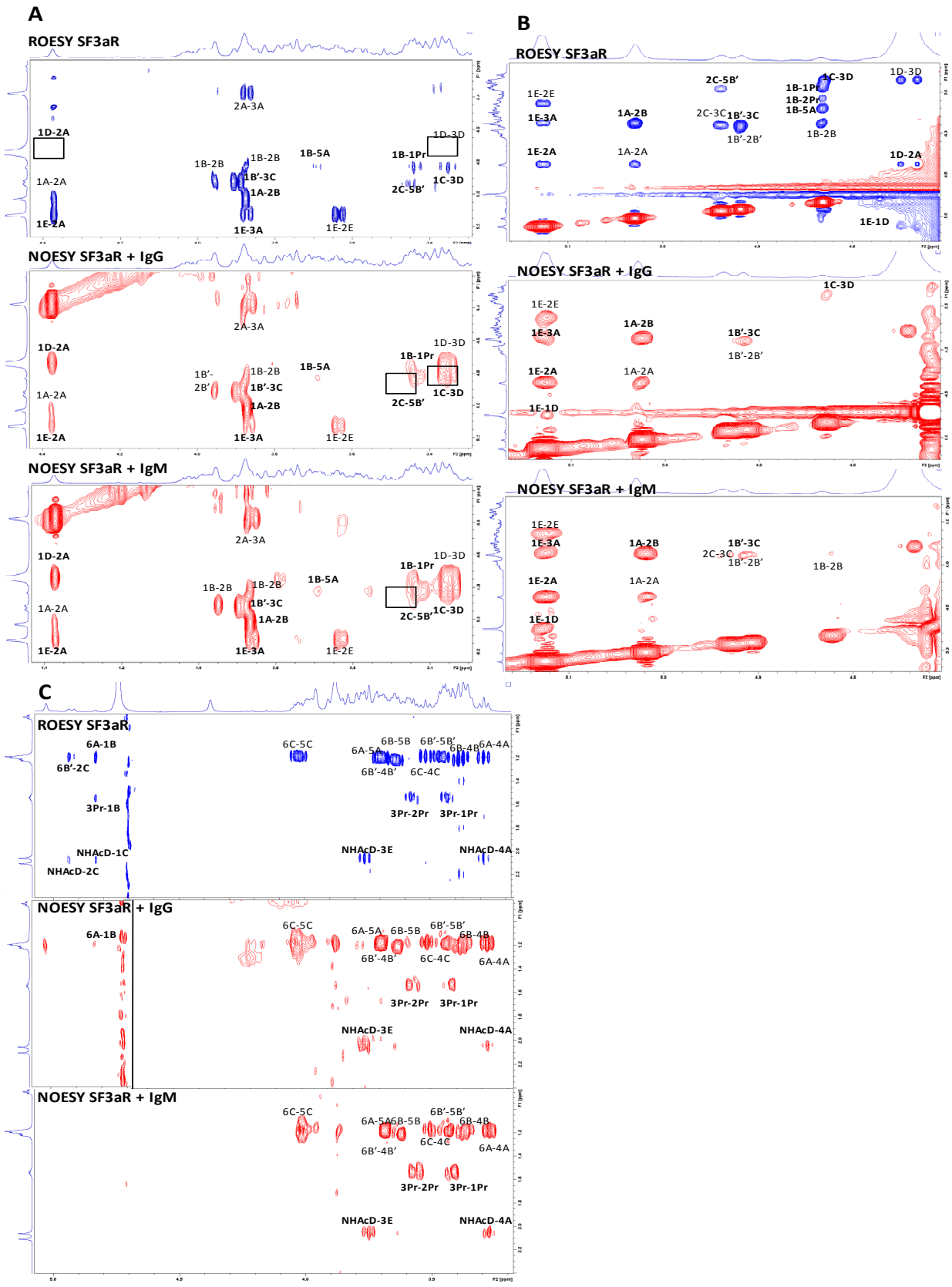


Figure 2.11 The key NOEs from relevant zones in the spectra found in the F2 (A, C) and F1 (B) dimension for SF3aR ligand. The ROESY spectra in the free state are shown always at the top (in blue) and the trNOESY in the presence of IgG (center) or IgM (below) in the trNOESY spectra are also shown.

The A-B linkage.-The A-B glycosidic torsion is present in most of the ligands (except SF3aO). The MD simulations predicted that the exoΦ/*syn*Ψ- geometry is predominant, although the alternative exoΦ/*syn*Ψ+ form is also present, independently of the single (with E) or double (with D and E) substitution at Rhap A. For the exoΦ/*syn*Ψ- geometry, a very strong H1A/H2B NOE (associated to a 2.2 Å distance) should be expected, much stronger than the H5A/H1B alternative (2.8 Å). Fittingly, the distances derived from ROESY spectra were in agreement with the presence of conformational equilibria, with a predominance of the exoΦ/*syn*Ψ- conformer in most cases. For example, in the case of ROESY spectrum of SF3aAA oligosaccharide (See Figure 2.13), the strong interresidual H1A'-H2B cross peak is observed, much stronger than the intra-residual H1A'-H2A'. The obtained data suggest the presence of a major Ψ- conformer.

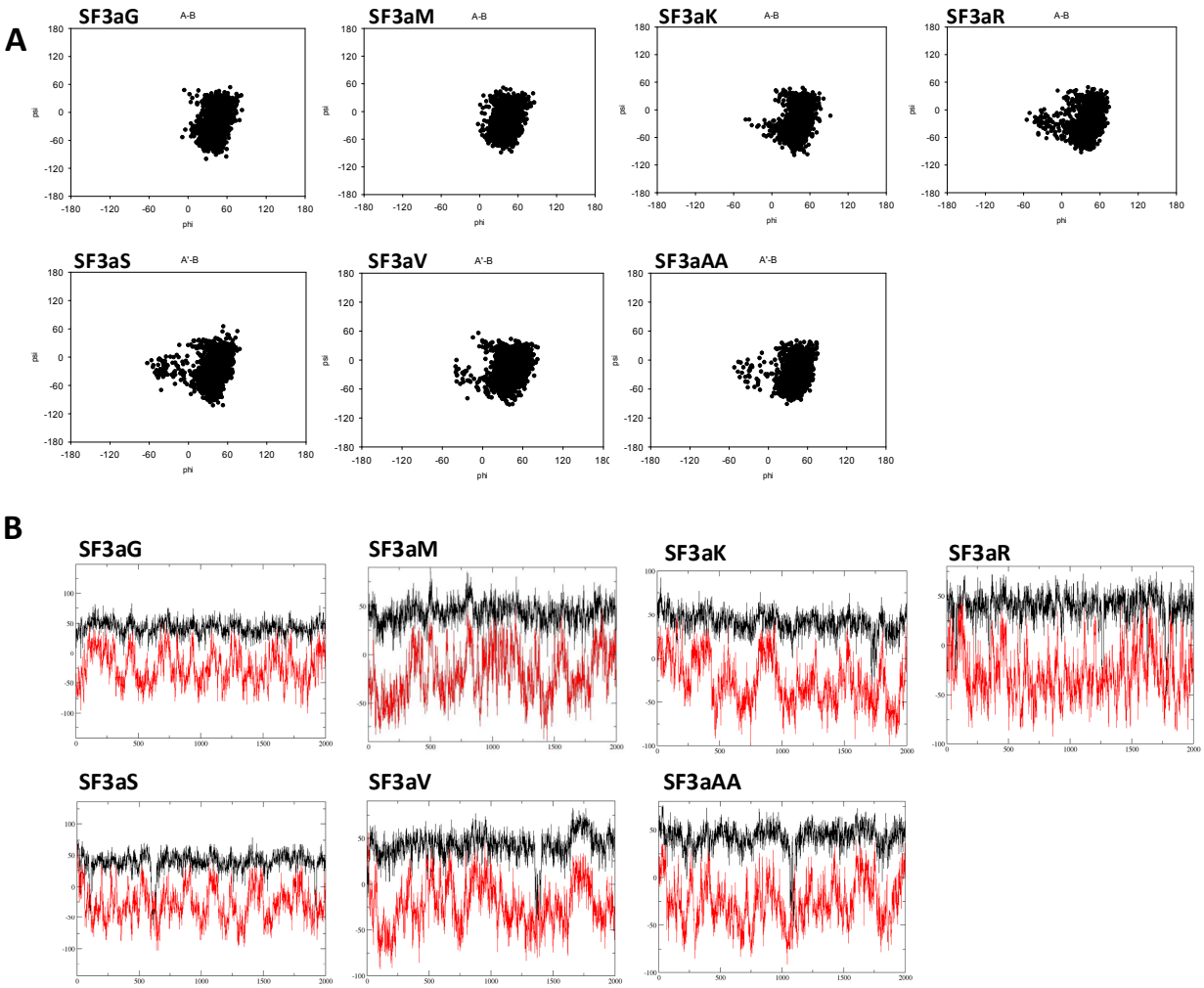


Figure 2.12 Plots of the available conformational space for the Φ/Ψ glycosidic linkages corresponding to the A/B linkage. The presence of a conformational equilibrium may be expected.

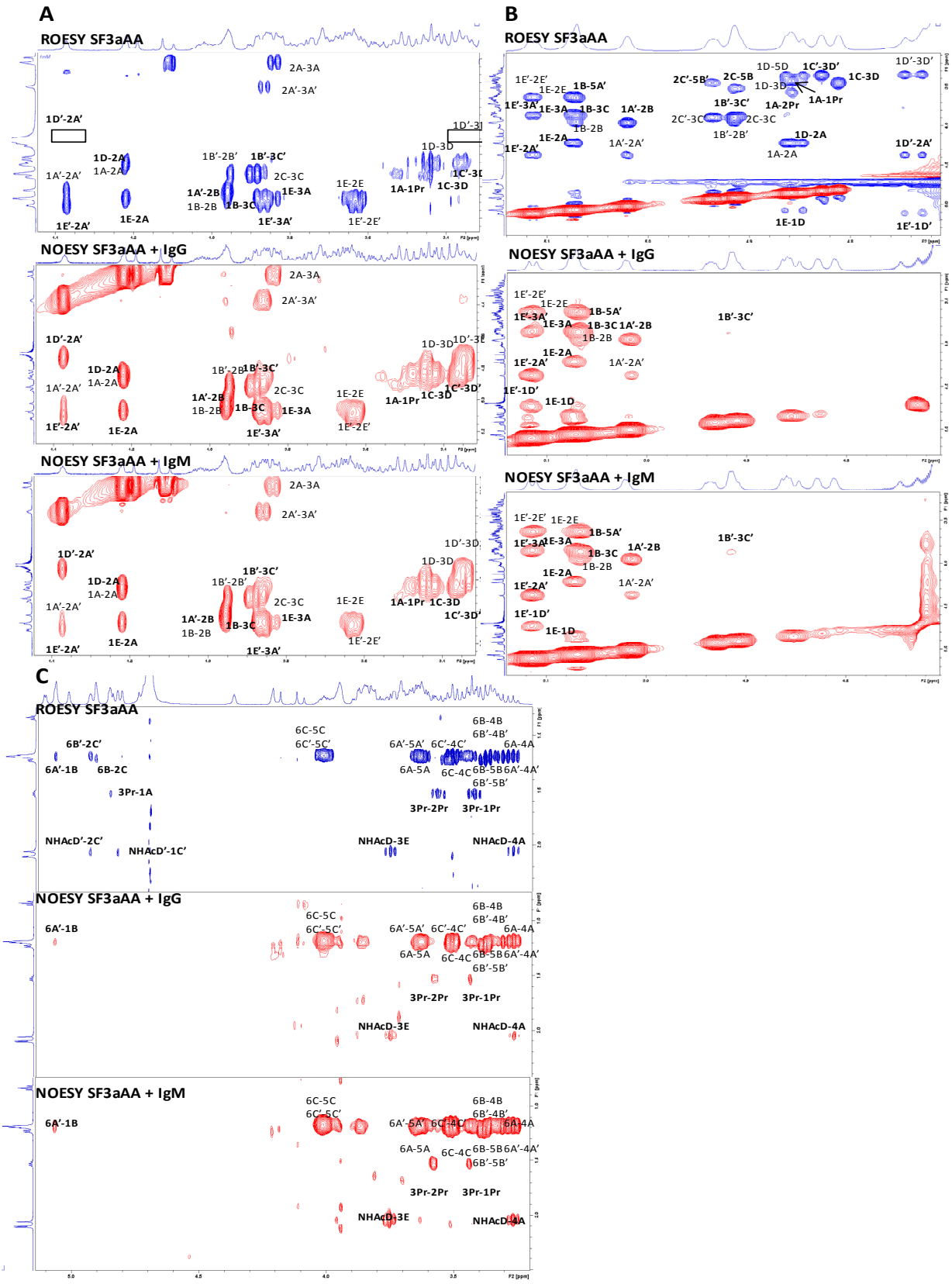


Figure 2.13 The key NOEs from relevant zones in the spectra found in the F2 (A, C) and F1 (B) dimension for SF3aAA ligand. The ROESY spectra in the free state are shown always at the top (in blue) and the trNOESY in the presence of IgG (center) or IgM (below) in the trNOESY spectra are also shown.

2.4 The bound state

2.4.1 The recognition of the designed oligosaccharides by IgG and IgM; *STD experiments*

The decoration of the linear E-A-B-C polysaccharide backbone with E and OAc residues provide different possibilities of epitope presentation. Moreover, the oligosaccharides studied herein were designed to provide evidences of the individual importance of the different residues, depending on its position (reducing, non reducing or internal) at the oligosaccharide sequence and of their relative relevance for the molecular recognition process. Moreover, for the longer oligosaccharides, the possibility of presenting an extended epitope could also be explored.

The simpler SF3aG oligosaccharide is a tetrasaccharide with residue E at the non reducing end and the acetylated C residue at the reducing end, further decorated with a n-propyl chain. The STD experiments provided clear evidences of the direct involvement of residue C, followed by the neighbor B residue (See Fig. 2.14). The acetylated residue is the key point for the interaction with both antibodies.

Thus, STD experiments were performed to study the relevance of residue GlcNAc D in the molecular recognition of the different oligosaccharides by IgG and IgM. While residue D is absent in SF3aG (EABC), there are two possible minimum "natural" repeating units of *S. flexneri* 3a serotype, either SF3aM (EABCD), with D at the reducing end, or SF3aK, with D at the non-reducing end (Fig. 2.2B). As mentioned above, even in the absence of residue D (SF3aG, EABC), the inspection of the STD spectra either in the presence of IgG or IgM evidenced the existence of interaction. Interestingly, the STD signal for Rha*p*A methyl group was clearly smaller than those of B and C in the presence of IgG (Fig. 2.15A). In the presence of IgM, the STD signals of H1 and H2 protons of Rha*p*C (Fig. 2.15B) were significantly stronger than those of the other residues.

Thus, the analysis of the STD intensities permitted to suggest that the BC fragment is the major epitope for IgG, while IgM especially recognizes residue C. When GlcNAc*p*D is located at the reducing end, as in SF3aM (EABCD), the strongest STD effects are those from the Rha*p*C, followed by the GlcNAc*p*D residues. Nevertheless, the inspection of the STD signals for the methyl moieties of the OAc and NAc groups suggested the existence of a better interaction of C than D (See Fig. 2.15) with IgG, while was fairly similar for IgM. The interaction of the E-A segment is less important.

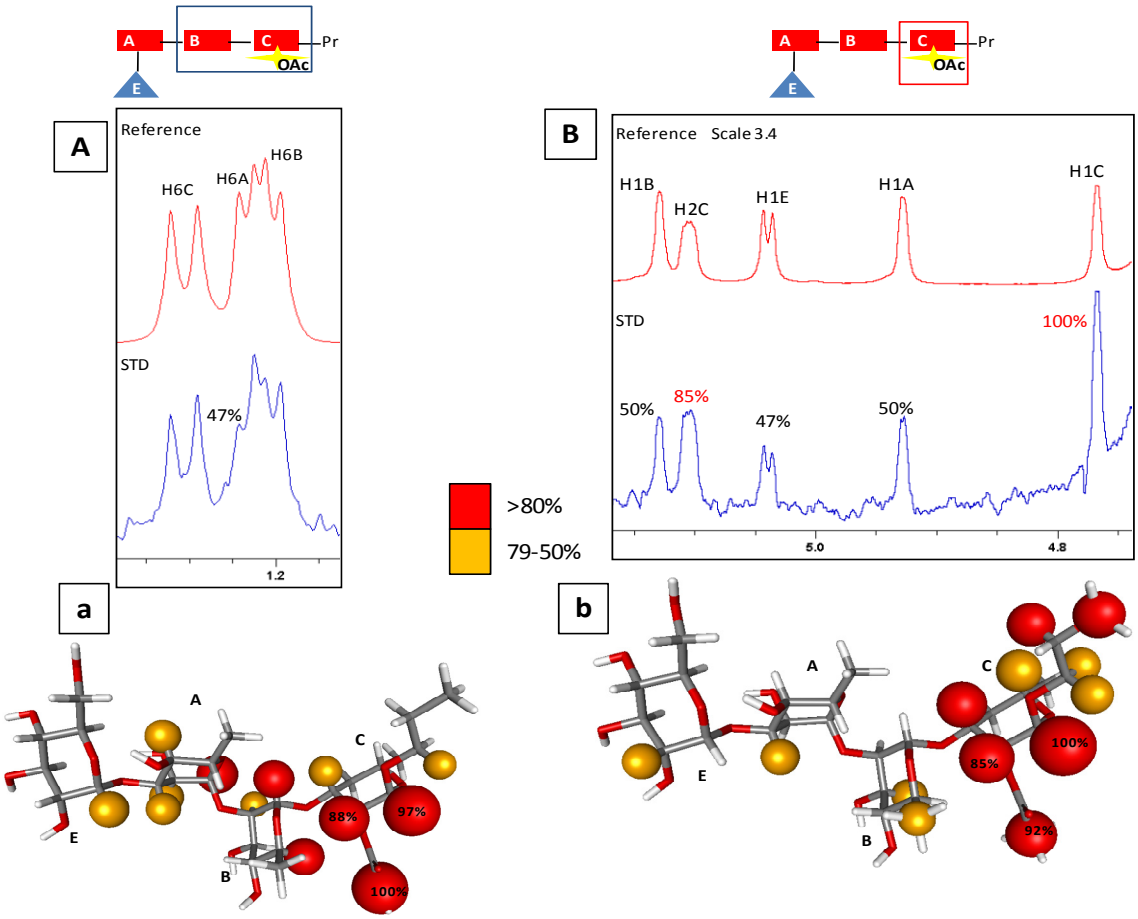


Figure 2.14 The interaction of SF3aG with IgG (A,a) and IgM (B,b) antibodies, as deduced from the STD experiments. The ligand: antibody molar ratio was 200:1 key. The observed STDs indicate the major involvement of the acetylated residue C, followed by its neighbors towards the non-reducing moiety.

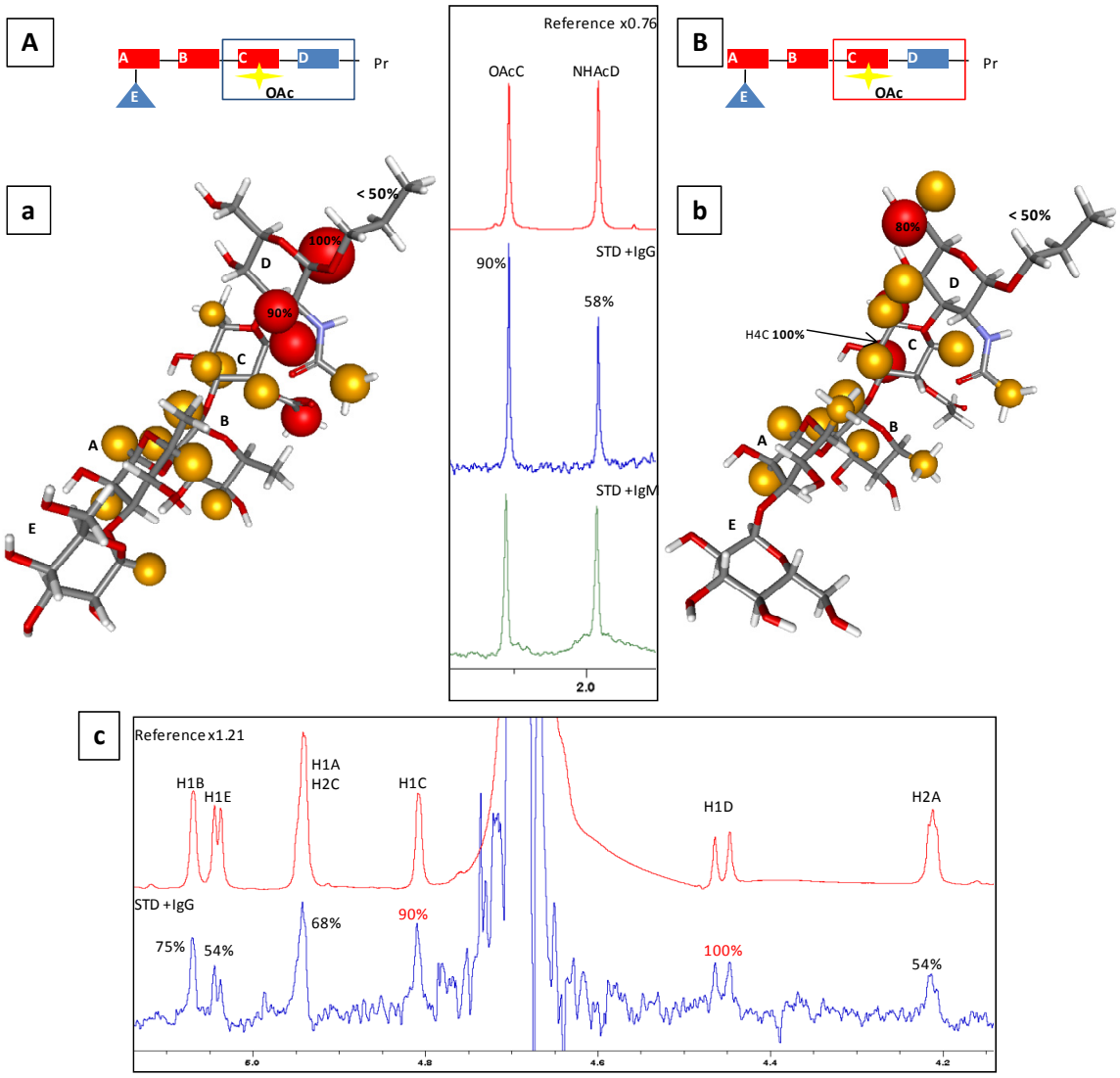


Figure 2.15 The interaction of SF3aM with IgG (A, a, c) and IgM (B, b) antibodies, as deduced from the STD experiments. The ligand: antibody molar ratio was 200:1 key. The observed STDs indicate the major involvement of the acetylated residue C, followed by residue D.

However, when the GlcNAc*p*D moiety is at the non reducing end, as in SF3aK (D(E)ABC), the interaction of D with the antibodies was significantly weaker, especially with IgG (Fig. 2.16). These results strongly suggest that the relative position of GlcNAc D in the oligosaccharide sequence is essential for making contacts with the antibody. The inspection of the STD intensities of the methyl groups of the A, B, and C Rha*p* units also implied a different involvement of these residues with the antibody. Clearly, residue A is less involved in the interaction than B and C, which is heavily involved.

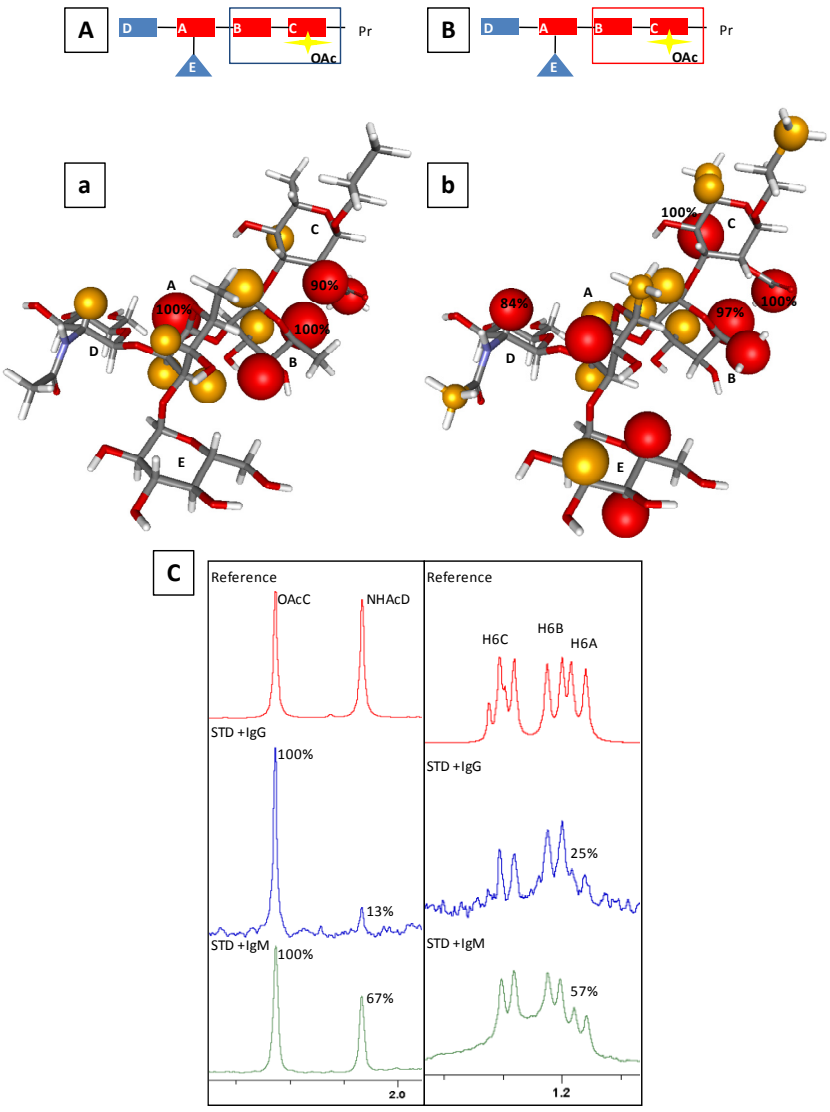


Figure 2.16 The interaction of SF3aK with IgG (A, a, C(blue)) and IgM (B, b, C(green)) antibodies, as deduced from the STD experiments. The ligand: antibody molar ratio was 200:1 key. The observed STDs indicate the major involvement of the acetylated residue C, followed by residue B. D is now not directly involved in binding

The E-A branch (Glc*p***E**α(1-3)Rha*p***A**)is commonly detected in all *S. flexneri* serotypes[13]. However, its relevance in *S. flexneri* serotype 3 for the immune recognition process by selective antibodies is already unknown. To prove the influence of the relative position of this substitution in the recognition of these oligosaccharides by IgG and IgM, STD-NMR experiments of penta and heptasaccharides displaying the E-A substitution at the reducing end, at the non-reducing end, and at both ends in the presence of IgG and IgM were performed (SF3aM, SF3aO, and SF3aS, respectively). The analysis of the STD experiments (Fig. 2.17) permitted to demonstrate that the three oligosaccharides were indeed recognized by both antibodies. In particular, in the

presence of IgG, (Fig. 2.17) Rha*p*C and surrounding B and D residues always displayed the strongest STD intensities, implying that is the key point of the interaction. Further analysis of the STD spectra of SF3aM and SF3aO systematically showed lower intensities of protons from E and A residues, indicating that the relative position of E-A fragment at either end of the pentasaccharide sequence is not important for making contacts with the antibodies. However, if the E-A branch is at both ends (SF3aS), the NMR experiments showed that the E'-A' branch at the non-reducing end provided significantly larger STD intensities than its E-A counterpart at the reducing end (Fig. 2.17), although always smaller than those for the B-C-D triad. Similar results were obtained in the presence of IgM (See Fig 2.18).

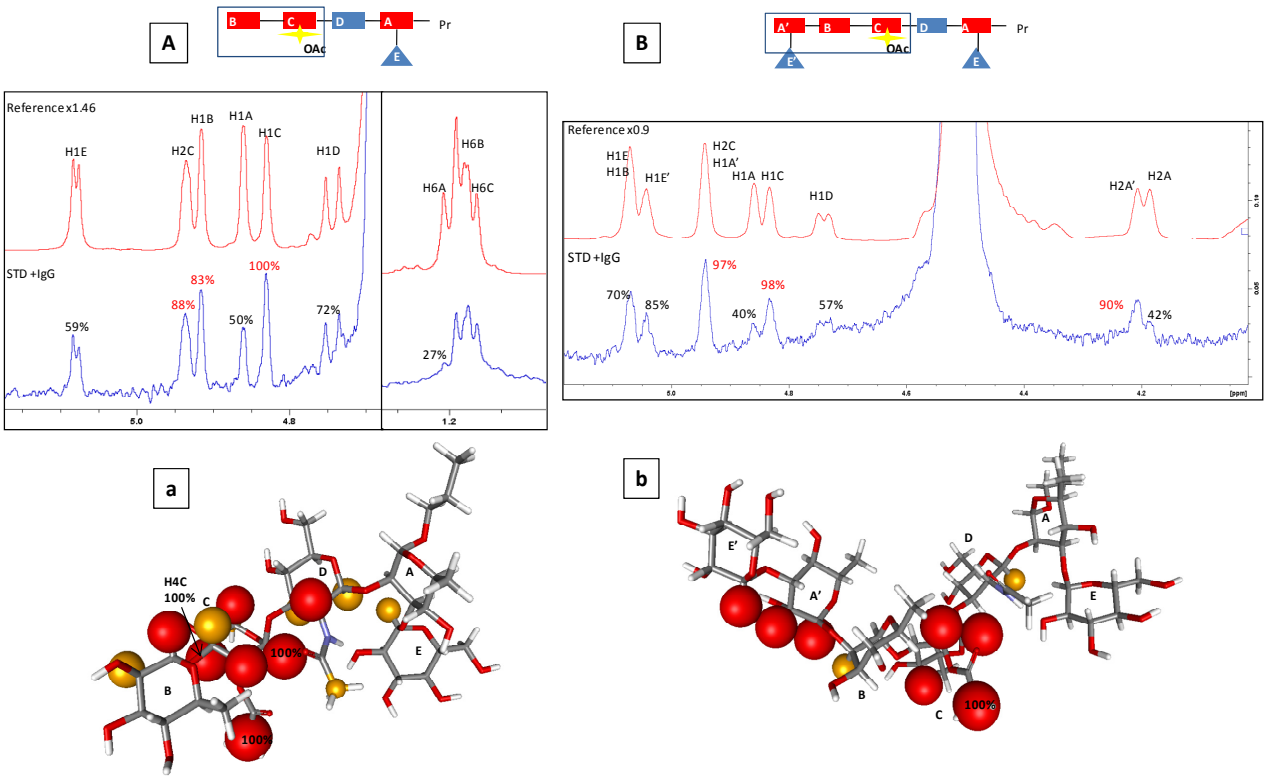


Figure 2.17 The interaction of SF3aO (left, A, a) and SF3aS (right, B, b) with IgG antibody, as deduced from the STD experiments. The ligand: antibody molar ratio was 200:1 key. The observed STDs indicate the major involvement of the acetylated residue C, followed by residue B. The reducing moieties are less involved in the binding event.

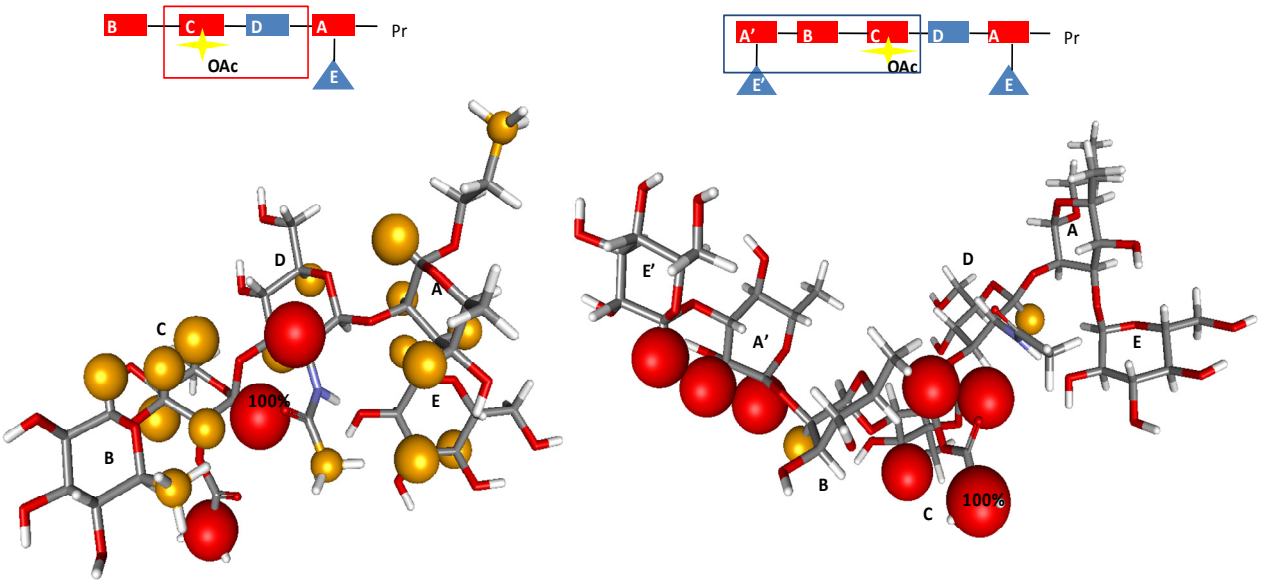


Figure 2.18 The interaction of SF3aO (left) and SF3aS (right) with IgM antibody, as deduced from the STD experiments.

The roles of a Rha*p*B substitution of SF3aO at the reducing end to provide hexasaccharide SF3aR (B'CD(E)A**B**) and that of a GlcNAc*p*D substitution of SF3aS at the non reducing end to produce octasaccharide SF3aV (**D'**(E)ABCD(E)A) were also tested in the presence of IgG and IgM. STD experiments evidenced the existence of the molecular recognition process with both antibodies, and permitted to define the epitope (See Fig. 2.19 and 2.20). For SF3aR in the presence of IgG, Rha*p*C and Rha*p*B residues are the key recognized epitope, followed by residue D (Fig. 2.19). Indeed, the observed STD intensities monotonically decrease from the B-C non-reducing end disaccharide moiety towards the reducing end. Similar results were detected for octasaccharide SF3aV in the presence of IgG (Fig. 2.20). Again, especially Rha*p*B and Rha*p*C, followed by *GlcNAcp* D, obtained the highest STD intensities. Interestingly, the STD effect at GlcNAc*p* D' was smaller than that at GlcNAc*p* D.

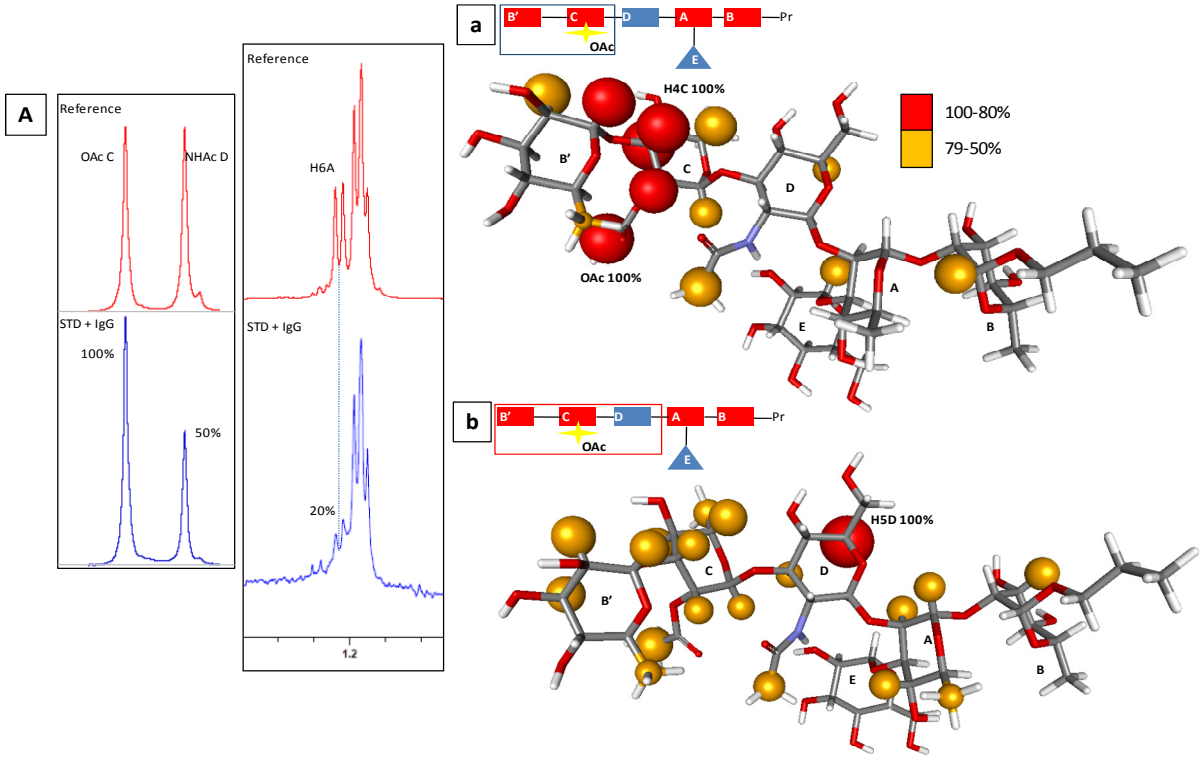


Figure 2.19 The interaction of SF3aR with IgG (A, a) and with IgM (b) antibodies, as deduced from the STD experiments. The ligand: antibody molar ratio was 200:1 key. The observed STDs indicate the major involvement of the acetylated residue C, followed by its flanking B and D residues. The reducing moieties are less involved in the binding event. The role of A is very minor. A similar behavior was observed for IgM binding.

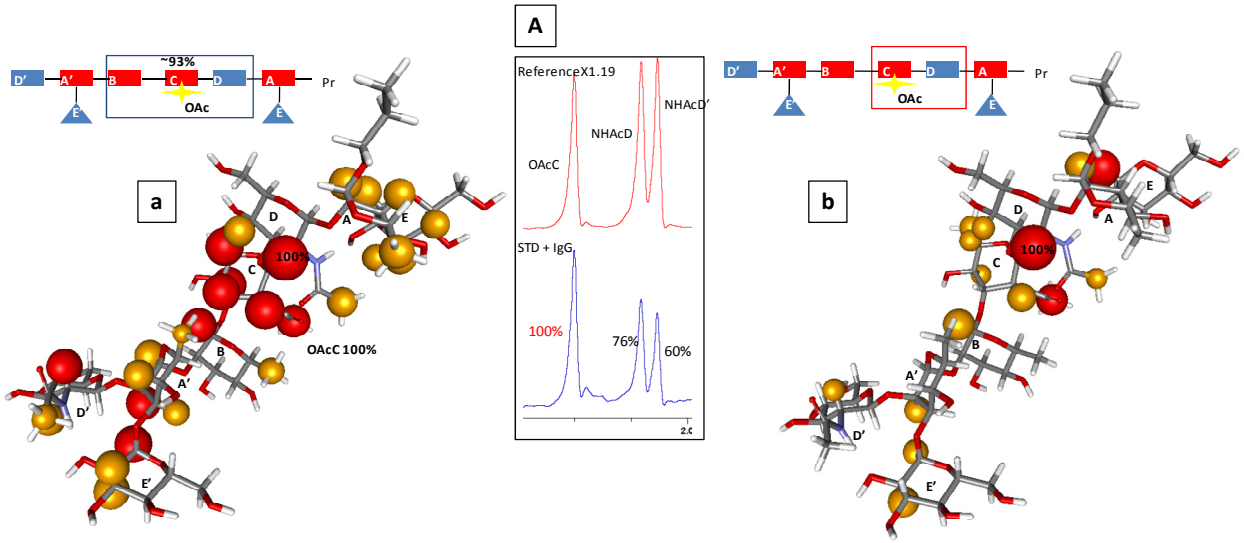


Figure 2.20 The interaction of SF3aV with IgG (A, a) and IgM (b) antibodies, as deduced from the STD experiments. The ligand: antibody molar ratio was 200:1 key. The observed STDs indicate the major involvement of the acetylated residue C, followed by its flanking B and D residues. The reducing moieties are less involved in the binding event. D is more involved than D' at the non-reducing end. A similar behavior was observed for IgM binding.

Finally, a long octasaccharide constituted by two repeating units of the *S. flexneri* serotype 3 (SF3aAA) was also employed. The STD results in the presence of IgG clearly indicated that the residues directly implicated in the recognition were Rha*p*C' at the "non-reducing" side of the octasaccharide, followed by GlcNAc*p*D' and Rha*p*A'. Again, the STD intensities decreased from the non-reducing towards the reducing end. Fittingly (Fig. 2.21A and a), STD signals for the methyl moieties of OAcC', OAcC and NAcD' demonstrate this. Similar results were obtained in the presence of IgM (See Fig. 2.21b).

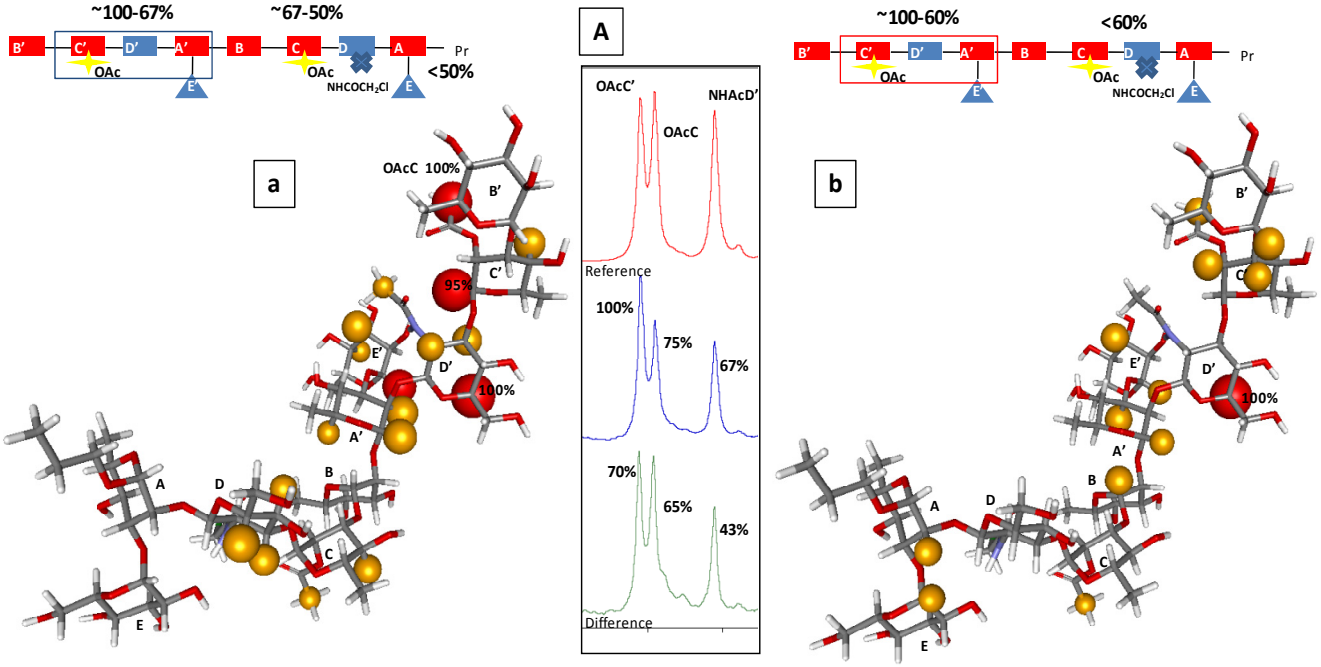


Figure 2.21 The interaction of SF3aAA with IgG (A(blue),a) and IgM (A(green),b) antibodies, as deduced from the STD experiments. The ligand: antibody molar ratio was 200:1 key. The observed STDs indicate the major involvement of the acetylated residue C', followed by its neighbor D' residue. The reducing moieties are less involved in the binding event. C' is more involved than C at the reducing end. A similar behavior was observed for IgM binding.

Thus, the STD results indicate that the acetylated residue C, flanked by B and D moieties, constitutes the major binding epitope for both antibodies, independently of its position at the oligosaccharide sequence. Nevertheless, for a double B-C-D substitution (as in SF3aAA), the major interaction takes place at the "non reducing" end. Residue A, as well as the reducing end propel moiety provide minor contacts with the antibodies, while the role of E, being minor, depends on its position in the sequence, becoming somehow more relevant for longer oligosaccharides.

2.4.2 The conformation in the bound state; *trNOESY experiments*

trNOESY experiments were also performed to gain additional structural information on the binding mode of the different oligosaccharides, to deduce the possible presence of a conformational selection process and to support the STD-based results. Fittingly, NOESY experiments (200ms mixing time) for the different oligosaccharides in their free state in water solution acquired at 600MHz and 298K did not give any NOE cross peaks for the sugar protons, thus indicating that the rotational global motion correlation time is around 200-300 microseconds timescale and the zero-NOE condition is fulfilled. In contrast, in the presence of either IgG or IgM, clear trNOESY negative cross peaks for all the oligosaccharides were observed, indicating the presence of a molecular recognition process. As already described, the existence of negative cross peaks is an unequivocal indication of the existence of increased effective rotational motion correlation times of the saccharide, accounting for the existence of an antibody-bound state. Thus, the analysis of these cross peaks permitted to deduce distance information that, in principle, could be related to the bound conformation of the saccharine. Therefore, the experimental key proton-proton distances across the glycosidic linkages were deduced and compared to those previously obtained in the free state in solution using ROESY measurements. Interestingly, for several molecules, the analysis of the cross peaks (Tables 2.1 A-H), suggested the existence of a particular conformational distribution around different glycosidic linkages, in some cases different to that existing in the free state. These facts indicate that the interaction of certain oligosaccharides with IgG and IgM antibodies modifies the oligosaccharide conformational behavior.

All the obtained data are gathered and summarized in Tables 2.1 A-H, although some comments may be added here to assess the goodness of the data and its agreement with the STD results. For instance, for the SF3aM/IgG complex (Fig. 2.22), the hydrogens at the propyl group clearly showed positive NOEs (i.e., H2-H1a and H2-H1b), while the rest of the sugar molecule showed negative NOES (i.e., H6B-H4B). This fact unequivocally indicates that the propel group does not directly interact with the IgG antibody and that, therefore, it does not take part in the interacting epitope. In fact, this fact was also evidenced by the STD data described above. In other cases, no clear negative cross peaks were observed for some of the monosaccharide, also

suggesting that these moieties did not maintain important contacts with the antibody. Similar data were extracted from the analysis of the complexes with IgM.

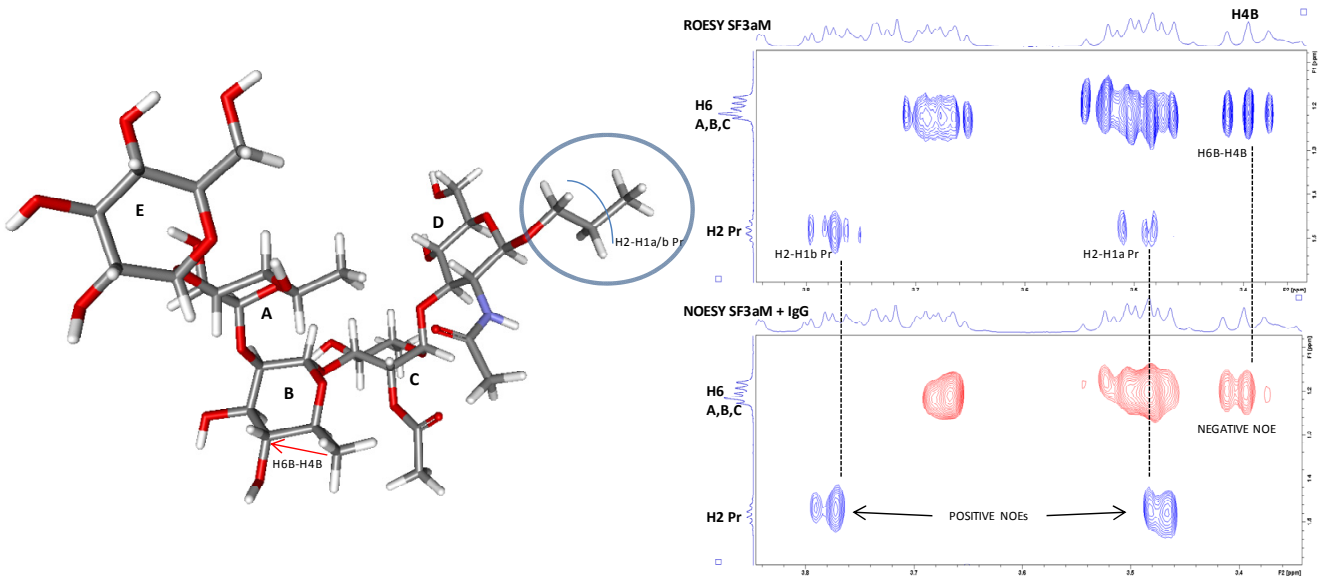


Figure 2.22 Region of propyl protons of SF3aM free ROESY spectrum and trNOESY of SF3aM /IgG complex.

The E-A linkage.-As described above, the conformational behavior of this glycosidic linkage maybe described by a conformational equilibrium, in which the exoΦ/*syn*Ψ- conformation is predominant in free solution. For SF3aG, weak or very weak interresidual H1E-H3A and H1E-H2A were obtained, suggesting that this region of the saccharide is not deeply involved in the binding (as also observed in the STD experiments). For SF3aM, a similar behavior was observed in the presence of IgG, while for IgM, the data suggested the recognition of the exoΦ/*syn*Ψ- geometry. This was also the case for SF3aK, SF3aO, and SF3aR when bound to the two different antibodies. For SF3aS (See Fig. 2.9) and SF3aV (Figure 2.24), which bears two different E-A linkages, negative NOE cross peaks were much stronger for the internal moiety than for the non-reducing terminal end. This fact strongly supports that the terminal and internal units are differently involved in the recognition process, with special interaction for the internal ones. The NOE pattern suggested that the internal units are recognized in with a major exoΦ/*syn*Ψ- geometry. For SF3aAA, the two E-A fragments are internal. The NOE patterns suggest a major recognition of the exoΦ/*syn*Ψ- conformation.

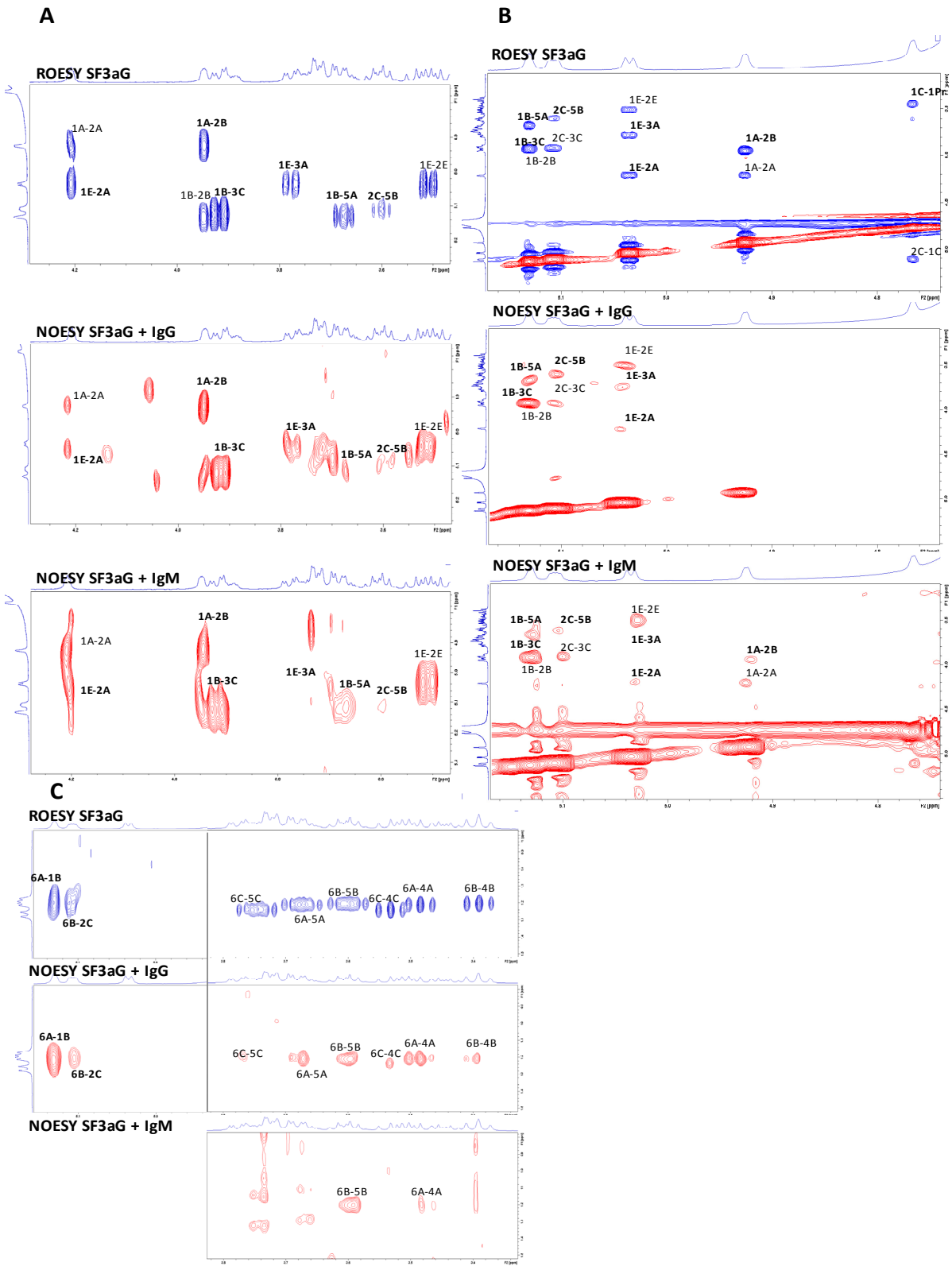


Figure 2.23 The key NOEs found for the SF3aG oligosaccharide found in the F2 (A, C) and F1 (B) dimension. The ROESY spectra in the free state are shown always at the top (in blue) and the trNOESY in the presence of IgG (center) or IgM (below) in the trNOESY spectra are also shown. Weak and very weak interresidual H1E-H2A and H1E-H3A cross peaks are observed. The obtained data suggest that the E-A fragment is not involved in the binding process.

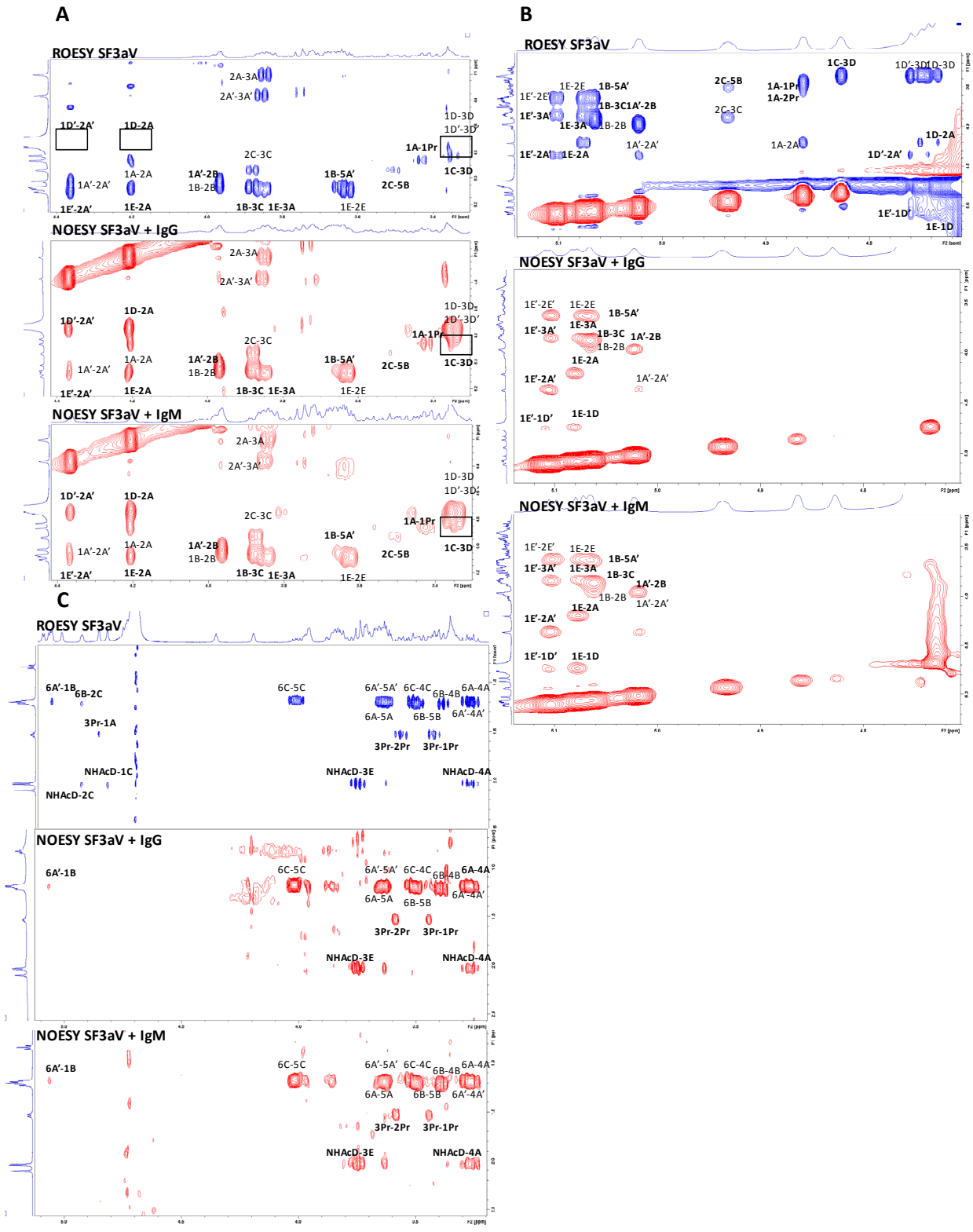


Figure 2.24 The key NOEs found for the SF3aV oligosaccharide in the F2 (A, C) and F1 (B) dimension. The ROESY spectra in the free state are shown always at the top (in blue) and the trNOESY in the presence of IgG (center) or IgM (below) in the trNOESY spectra are also shown.

The D-A linkage.-Regarding the D-A(E) glycosidic linkage, the proximity of H1D to the water signal precluded the proper quantification of the interesidual H1D-H2A cross peak with respect to the intraresidual H1D-H3D. However, the strong intensity of the H1D-H2A cross peak was correlated with the existence of a major exoΦ/*syn*Ψ- geometry for the D-A linkage of SF3aK, SF3aO, SF3aR, and SF3aS. For SF3aK, the H1D-H2A cross peak was significantly smaller than the H1B-H3C analogue, indicating the different involvement of the two glycosidic linkages in the binding event (See Fig 2.25). The D-A linkage has less importance for the recognition, especially for IgG, as also observed in the STD experiments. For the two D-A glycosidic linkages of SF3aAA (See Fig. 2.13) and especially of SF3aV (Fig. 2.24), the cross peak pattern intensities suggested the major involvement of the internal linkages than the reducing end analogues, in close agreement with the STD results. In all cases, the exoΦ/*syn*Ψ- was the major bound conformer.

The B-C linkage.-For the B-C linkage, which could be defined as a major exoΦ/*syn*Ψ- geometry in the free state for all the oligosaccharides, the analysis of the trNOESY data pointed out to the same behavior when the different oligosaccharides were bound to either IgG or IgM. The bound state conformation to the two antibodies corresponds to the major one existing in solution. For example, in the case of SF3aR, The cross peak intensity of the interresidual H1B'-H3C is much stronger than the intraresidual one, indicating that the exoΦ/*syn*Ψ- geometry is the major one recognized by both antibodies (See Fig 2.11).

The C-D linkage.- For the C-D glycosidic torsion, the key H1C-H3D cross peaks was fairly intense in the bound state to IgG and IgM. These facts probably indicate that the exoΦ/*syn*Ψ+ region is the bound conformer to these antibodies. Again, this region also corresponds to the binding epitope of most of these molecules for the studied antibodies, as deduced from the STD experiments. For example, in the case of SF3aM (See Figure 2.5B) and SF3aV (See Figure 2.24B), the cross peak intensity of the interresidual H1C-H3D is fairly strong, indicating that the exoΦ/*syn*Ψ+ geometry is the major one recognized by both antibodies in both cases.

Figure 2.25 The key NOEs found for the SF3aK oligosaccharide in the F2 (A, C) and F1 (B) dimension. The ROESY spectra in the free state are shown always at the top (in blue) and the trNOESY in the presence of IgG (center) or IgM (below) in the trNOESY spectra are also shown

The A-B linkage. -The A-B linkage showed basically identical behavior in the free and bound states to both antibodies, with a clear predominance of the exoΦ/*syn*Ψ-geometry, as deduced from the intensity of the interresidual H1A'-H2B cross peak, much stronger than the corresponding intraresidual one. For example, in the case of SF3aM (See Figure 2.5 B) and SF3aR (See Figure 2.11B), the cross peak intensities of the interresidual H1A-H2B is fairly strong, significantly stronger than the intraresidual H1A-H2A, indicating that the exoΦ/*syn*Ψ+ geometry is the major one recognized by both antibodies. Also in the case of SF3aAA (See Fig. 2.13B), the cross peak intensities of the interresidual H1A'-H2B is fairly strong, significantly stronger than the intraresidual H1A'-H2A', indicating that the exoΦ/*syn*Ψ+ geometry is the major one recognized by both antibodies.

2.5 Conclusions

The NMR analysis has shown that there is major recognition epitope in the studied oligosaccharides and they display a major conformation in their bound states to the antibodies, through a conformational selection process. In the absence of a three-dimensional model of the antibodies, no clear-cut conclusions may be extracted on the driving forces for the interaction to take place. In any case, we decided to explore the occurrence of different polar/non-polar surfaces of the different ligands using the previously calculated bound conformation. The corresponding data are given in Figures 2.26A-C. In the case of Fig 2.26A, the polar regions for each oligosaccharide were highlighted in red. The analysis of the computed molecular surfaces showed that the major polar zones correspond to the non interacting zone of the ligand, thus supporting the STD-NMR results. On the other hand, the non-polar regions for the same oligosaccharides were highlighted in grey in Figs. 2.26B and C. Analysis of molecular surfaces showed that the major non-polar zones correspond to the closest zone of the ligand to the antibody surface according to the STD-NMR results previously described.

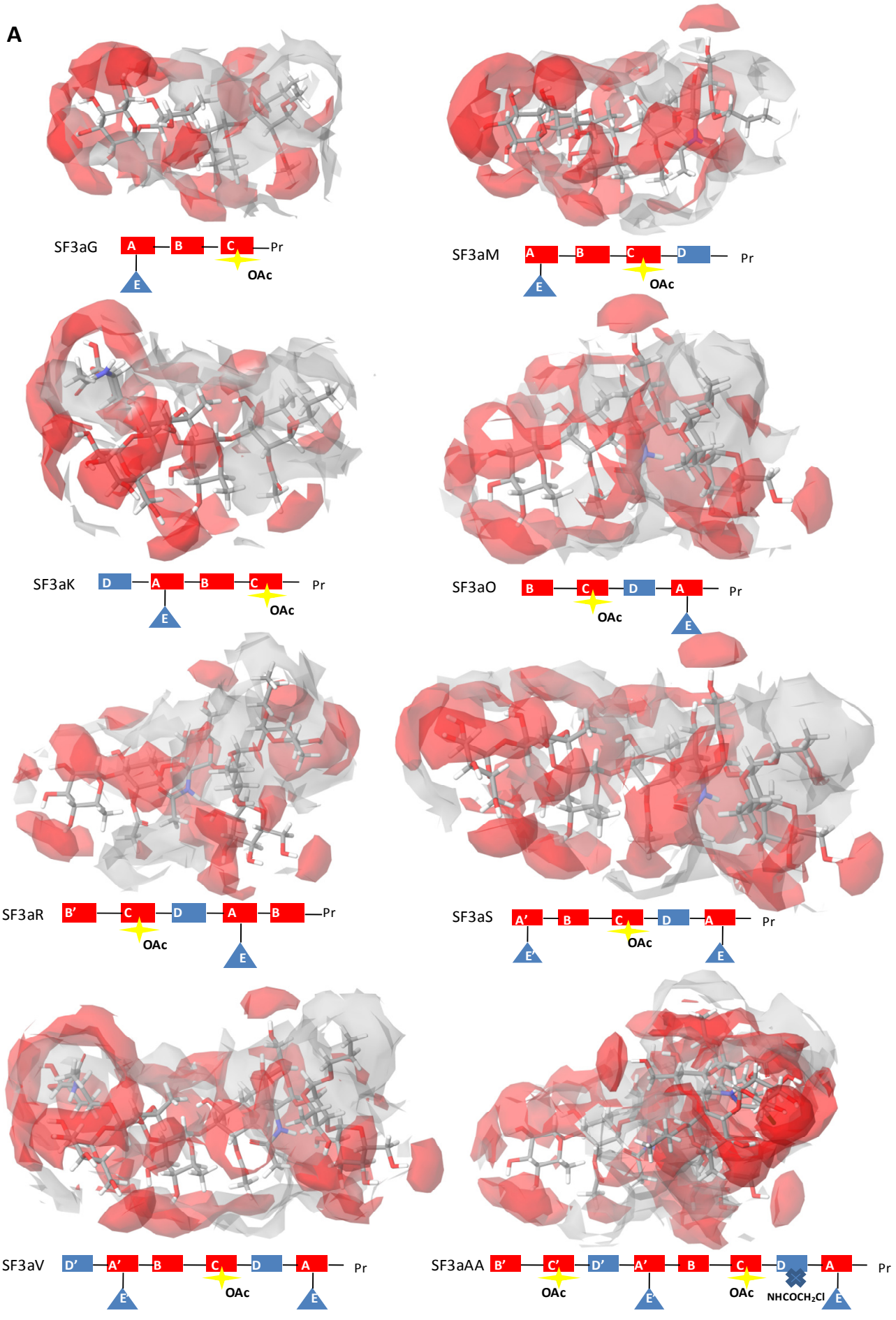
As example, for SF3aM, upon inspection of the polar/non-polar molecular surfaces, it can be deduced that the E and A residues display a polar surface. This region should be solvent exposed (See SF3aM ligand, Fig 2.26A). In fact, for this SF3aM

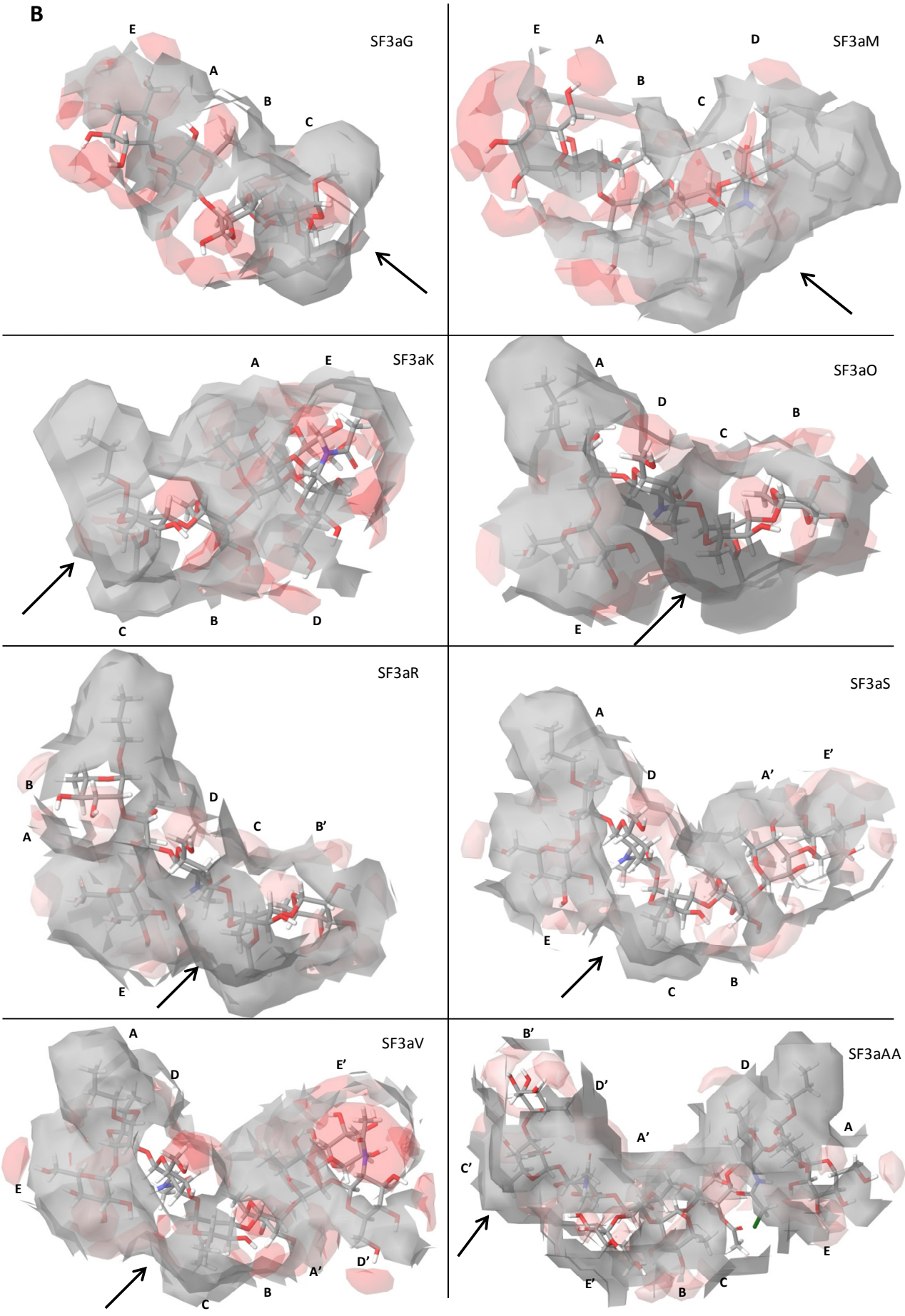
ligand (Figs 2.26B and C), the non-polar region, in grey, contains C and D residues. These moieties are making the key contacts with the antibody according to the NMR experiments. This fact implies that this region of the structure is making non-polar interactions with the protein surface. Similar observations can be derived for the rest of the oligosaccharide/mAb systems.

In summary, we have demonstrated that the combination of STD and trNOESY NMR and MD simulations could provide important information about the molecular recognition of synthetic oligosaccharides towards two specific antibodies IgG C7-37 and IgM G19-2. Firstly, ROESY-NMR experiments gave us structural information about the major Φ/Ψ values of each glycosidic linkage and therefore, the existing conformational equilibrium.

Secondly, our study has provided experimental evidences on the individual importance of the different residues, depending on its position at the oligosaccharide sequence, also for longer oligosaccharides, and their relative relevance for the molecular recognition process. The information derived from STD-NMR experiments clearly showed evidences of the direct involvement of residue C followed by its neighbor residues (Rha*p*B-**Rha*p*C**-GlcNAc*p*D) in every synthetic oligosaccharide in the presence of both mAbs. The detailed analysis of trNOESY-NMR experiments in the presence of protective antibodies IgG and IgM, gave insights into the presence of a conformational selection process, always supported by the STD-based results. The polar/non-polar surface generation of all ligand-bound conformers supported the results obtained from the NMR experiments (See Fig 2.26A-C). In all instances case, the region directly involved in the interaction according to STD experiments (Residue C and surroundings; Rha*p*B-**Rha*p*C**-GlcNAc*p*D) always coincided with a clear non-polar area in the bound oligosaccharide three-dimensional structure (See arrows in Fig 2.26B and Fig 2.26C).

A


SF3aG
A
B
C
Pr
E
OAc
SF3aM
A
B
C
D
Pr
E
OAc
SF3aK
D
A
B
C
Pr
E
OAc
SF3aO
B
C
D
A
Pr
OAc
E
SF3aR
B'
C
D
A
B
Pr
OAc
E
SF3aS
A'
B
C
D
A
Pr
E'
OAc
E
SF3aV
D'
A'
B
C
D
A
Pr
E'
OAc
E
SF3aAA
B'
C'
D'
A'
B
C
D
A
Pr
OAc
E'
OAc
NHCOCH2Cl
E


B
SF3aG
E
A
B
C
SF3aM
E
A
D
B
C
SF3aK
A
E
C
B
D
SF3aO
A
D
C
B
E
SF3aR
B
A
D
C
B'
E
SF3aS
A
D
E'
A'
E
C
B
SF3aV
A
D
E'
E
A'
D'
B
C
SF3aAA
B'
D'
D
A'
A
C'
E
E'
B
C

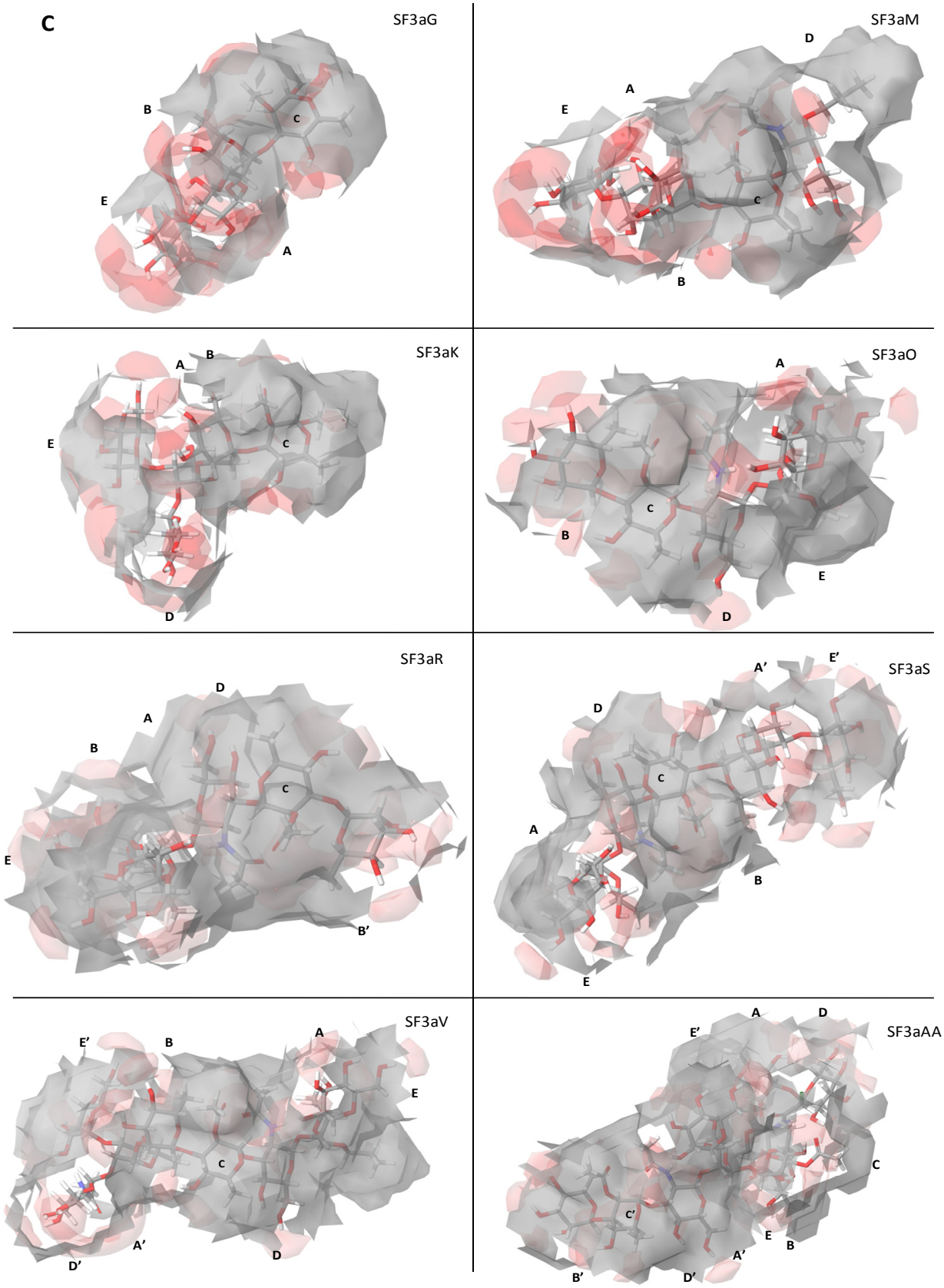


Figure 2.26 A-C Polar/Non polar surfaces generated from all ligand-bound conformers emphasizing in A polar zones in red and non-polar zones in grey in B and C. In Figure B the black arrows detail the epitope of each ligand observed in the STD experiments.

TABLE 2.1 A
SF3aG

	CONFORMATION 1 (syn-)	CONFORMATION 2 (syn+)	Experimental FREE STATE	Experimental IgG	Experimental IgM
	E-A				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcE H1-RhaA H2	2.1	4	2.4	very weak	very weak
GlcE H1-RhaA H3	2.7	2.3	2.5	very weak	very weak
	A-B				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaA H1-RhaB H2	2.2	2.5	2.2	2.1	2.4
RhaA H5-RhaB H1	2.7	2.3	2.6	>2.7	>2.7
RhaA CH36-RhaB H1	3.2	4	3.6	3.4	>3.5
	B-C				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaB H1-RhaC H3	2.1	2.6	2.1	2.1	2.1
RhaB H5-RhaC H2	2.5	2.4	2.7	2.4	2.5
RhaB CH36-RhaC H2	3.3	4	3.5	3.3	nd

TABLE 2.1 B
SF3aM

	CONFORMATION 1 (syn-)	CONFORMATION 2 (syn+)	Experimental FREE STATE	Experimental IgG	Experimental IgM
	E-A				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcE H1-RhaA H2	2.1	4	2.4	2.8	2.4
GlcE H1-RhaA H3	2.7	2.3	2.5	>3.0	2.7
	A-B				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaA H1-RhaB H2	2.2	2.5	2.3	2.1	2.2
RhaA H5-RhaB H1	2.7	2.3	2.6	2.6	2.6
RhaA CH36-RhaB H1	3.3	4	3.3	nd	3.3
	B-C				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaB H1-RhaC H3	2.1	2.6	2.2	2.1	2.1
RhaB H5-RhaC H2	2.5	2.4	2.6	2.4	2.5
RhaB CH36-RhaC H2	3.8	4	3.8	nd	nd
	C-D				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaC H1-GlcNAcD H3	2.3	2.5	2.5	2.5	2.5
RhaC H5-GlcNAcD H4	2.7	3.5	>3.3	nd	nd

TABLE 2.1 C
SF3aK

	CONFORMATION 1 (syn-)	CONFORMATION 2 (syn+)	Experimental FREE STATE	Experimental IgG	Experimental IgM
	E-A				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcE H1-RhaA H2	2.2	4	2.6	2.5	2.5
GlcE H1-RhaA H3	2.8	2.2	2.6	2.7	2.8
	D-A				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcNAcD H1-RhaA H2	2.3	2.9	2.4	2.6	2.5
GlcNAcD NHAc-RhaA H4	5.2	3.5	>4	>4	>4
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcNAcD H1-GlcE H1	2.8	5.3	3.2	3	2.9
GlcNAcD NHAc-GlcE H3	5.8	3.9	>4	>4	>4
	A-B				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaA H1-RhaB H2	2.2	2.6	2.3	2.2	2.2
RhaA CH36-RhaB H1	3.3	4.1	3.4	3.5	3.8
	B-C				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaB H1-RhaC H3	2.2	2.6	2.2	2.2	2.2
RhaB CH36-RhaC H2	3.3	4.2	nd	3.3	3.6

TABLE 2.1 D
SF3aO

	CONFORMATION 1 (syn-)	CONFORMATION 2 (syn+)	Experimental FREE STATE	Experimental IgG	Experimental IgM
	E-A				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcE H1-RhaA H2	2.2	4	2.4	2.3	2.4
GlcE H1-RhaA H3	2.8	2.2	2.5	2.9	2.7
	D-A				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcNAcD H1-RhaA H2	2.2	2.6	2.5	2.2	2.2
GlcNAcD NHAc-RhaA H4	5.2	3.5	>3.5	>4	>4
	D-E (Long range)				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcNAcD H1-GlcE H1	2.4	6.1	2.6	NOE	2.4
GlcNAcD NHAc-GlcE H3	5.8	3.6	>4	>4	>4
	B-C				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaB H1-RhaC H3	2.2	2.6	2.2	2.2	2.2
RhaB H5-RhaC H2	2.5	2.4	2.6	2.6	2.7
RhaB CH36-RhaC H2	3.2	4	3.8	nd	nd
	C-D				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaC H1-GlcNAcD H3	2.3	2.5	2.5	2.5	2.5
RhaC H5-GlcNAcD H4	2.7	3.5	nd	nd	nd

TABLE 2.1 E
SF3aR

	CONFORMATION 1 (syn-)	CONFORMATION 2 (syn+)	Experimental FREE STATE	Experimental IgG	Experimental IgM
	E-A				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcE H1-RhaA H2	2.2	4.1	2.5	2.5	2.4
GlcE H1-RhaA H3	2.8	2.25	2.7	2.7	2.6
	D-A				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcNAcD H1-RhaA H2	2.2	2.7	2.7	2.2	2.2
GlcNAcD NHAc-RhaA H4	4.6	3.5	>3.5	>4	>4
	D-E (Long range)				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcNAcD H1-GlcE H1	2.4	6	2.7	3	2.7
GlcNAcD NHAc-GlcE H3	6	3.7	4	>4	>4
	A-B				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaA H1-RhaB H2	2.2	2.5	2.2	2.2	2.2
RhaA H5-RhaB H1	2.6	2.5	2.5	2.5	2.7
RhaA CH36-RhaB H1	3.3	4.1	3.5	NOE	nd
	B'-C				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaB' H1-RhaC H3	2.2	2.7	2.2	2.2	2.3
RhaB' H5-RhaC H2	2.5	2.6	2.7	2.5	nd
RhaB' CH36-RhaC H2	3.2	4.4	3.6	nd	nd
	C-D				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaC H1-GlcNAcD H3	2.3	2.5	2.5	2.6	2.6
RhaC H5-GlcNAcD H4	2.6	3.5	nd	nd	nd

TABLE 2.1 F
SF3aS

	CONFORMATION 1 (syn-)	CONFORMATION 2 (syn+)	Experimental FREE STATE	Experimental IgG	Experimental IgM
	E'-A'				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcE' H1-RhaA' H2	2.2	4.1	2.5	>3	3
GlcE' H1-RhaA' H3	2.8	2.3	2.5	2.8	2.8
	A'-B				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaA' H1-RhaB H2	2.2	2.5	2.2	2.2	2.2
RhaA' CH36-RhaB H1	3.3	4.1	3.5	3.5	3.5
	B-C				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaB H1-RhaC H3	2.1	2.5	2.1	2.1	2.1
RhaB CH36-RhaC H2	3.4	3.9	3.7	nd	nd
	C-D				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaC H1-GlcNAcD H3	2.3	2.5	2.5	2.5	2.5
RhaC H5-GlcNAcD H4	2.5	3.5	nd	nd	nd
	D-A				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcNAcD H1-RhaA H2	2.2	2.7	2.5	2.5	2.5
GlcNAcD NHAc-RhaA H4	4.6	3.7	3.9	>4	>4
	E-A				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcE H1-RhaA H2	2.2	4.1	2.5	2.6	2.6
GlcE H1-RhaA H3	2.8	2.3	2.5	2.8	2.7
	D-E (Long range)				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcNAcD H1-GlcE H1	2.4	5	2.8	2.9	2.9
GlcNAcD NHAc-GlcE H3	5	3.6	4	>4	>4

TABLE 2.1 G
SF3aV

	CONFORMATION 1 (syn-)	CONFORMATION 2 (syn+)	Experimental FREE STATE	Experimental IgG	Experimental IgM
	E'-A'				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcE' H1-RhaA' H2	2.1	4.1	2.7	>2.7	>2.7
GlcE' H1-RhaA' H3	2.8	2.3	2.6	>2.8	>2.8
	D'-A'				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcNAcD' H1-RhaA' H2	2.2	2.7	2.6	2.2	2.2
	D'-E' (Long range)				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcNAcD' H1-GlcE' H1	2.4	4.5	>3.2	>3	>3
GlcNAcD' NHAc-GlcE' H3	5.7	3.7	>4	nd	nd
	A'-B				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaA' H1-RhaB H2	2.2	2.5	2.2	2.2	2.2
RhaA' CH36-RhaB H1	3.4	4	3.5	3.5	3.5
	B-C				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaB H1-RhaC H3	2.2	2.5	2.2	2.2	2.2
RhaB CH36-RhaC H2	3.3	4	3.3	nd	nd
	C-D				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaC H1-GlcNAcD H3	2.3	2.5	2.5	2.6	2.6
RhaC H5-GlcNAcD H4	2.7	3.5	nd	nd	nd
	D-A				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcNAcD H1-RhaA H2	2.2	2.5	2.6	2.4	2.4
	E-A				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcE H1-RhaA H2	2.1	4.1	2.5	2.6	2.6
GlcE H1-RhaA H3	2.8	2.2	2.5	2.8	2.8
	D-E (Long range)				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcNAcD H1-GlcE H1	2.4	4.7	2.9	2.9	2.9
GlcNAcD NHAc-GlcE H3	5.1	3.7	4	>4	>4

TABLE 2.1 H
SF3aAA

	CONFORMATION 1 (syn-)	CONFORMATION 2 (syn+)	Experimental FREE STATE	Experimental IgG	Experimental IgM
	B'-C'				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaB' H1-RhaC' H3	2.2	2.5	2.2	nd	nd
RhaB' H5-RhaC' H2	2.5	2.4	2.5	nd	nd
RhaB' CH36-RhaC' H2	3.2	4	3.7	nd	nd
	C'-D'				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaC' H1-GlcNAcD' H3	2.3	2.5	2.5	2.6	2.6
RhaC' H5-GlcNAcD' H4	2.7	3.5	nd	nd	nd
	D'-A'				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcNAcD'H1-RhaA' H2	2.2	2.5	2.5	2.2	2.2
GlcNAcD' NHAc-RhaA' H4	4.8	3.7	>4	>4	>4
	E'-A'				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcE' H1-RhaA' H2	2.2	4.1	2.6	2.4	2.4
GlcE' H1-RhaA' H3	2.8	2.3	2.6	2.6	2.6
	D'-E' (Long range)				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcNAcD' H1-GlcE' H1	2.4	4.7	>3	2.7	2.7
GlcNAcD' NHAc-GlcE' H3	3.6	5.2	>4	>4	3.5
	A'-B				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaA' H1-RhaB H2	2.3	2.4	2.2	2.2	2.2
RhaA' CH36-RhaB H1	3.4	3.9	3.5	3.5	3.5
	B-C				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaB H1-RhaC H3	2.2	2.5	2.2	2.2	2.2
RhaB H5-RhaC H2	2.5	2.4	2.5	nd	nd
RhaB CH36-RhaC H2	3.3	3.9	3.7	nd	nd
	C-D				
Distance	Calculated (Angs.)	Calculated (Angs.)			
RhaC H1-GlcNAcD H3	2.3	2.5	2.5	2.6	2.6
RhaC H5-GlcNAcD H4	2.7	3.5	nd	nd	nd
	D-A				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcNAcD H1-RhaA H2	2.2	2.5	nd	nd	nd
	E-A				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcE H1-RhaA H2	2.2	4	2.5	2.4	2.4
GlcE H1-RhaA H3	2.8	2.3	2.6	2.5	2.6
	D-E (Long range)				
Distance	Calculated (Angs.)	Calculated (Angs.)			
GlcNAcD H1-GlcE H1	2.4	4.6	3	2.7	2.7

Tables 2.1 A-H Inter-glycosidic distances of every SF3a oligosaccharide derived from MD simulations (first and second columns), derived from ROESY spectra (third column) and in the presence of the Immunoglobulins; derived from NOESY spectra in the presence of IgG (fourth column) and IgM (fifth column).

2.6 Materials and Methods

Nomenclature

The glycosidic torsion angles were defined employing the NMR-based convention, referred to the glycosidic and aglyconic hydrogens. In particular, φ was defined as H1'-C1'-O1-CX and ψ as HX-CX-O1-C1'. X refers to either position 2 or 3 for the (1→2) and (1→3) linkages, respectively.

Methods

MD simulations of SF3a oligosaccharides

The initial geometries of the SF3a ligands were generated by using the LEAP module of AMBER10[14]. Structures were first minimized and then the solute was placed in a rectangular TIP3P box with water molecules, occupying up to 11Å from the sugar center of mass. The MD simulations were run for 4ns using GLYCAM06 force field[15], started after four initial minimization steps with the solute restrained in its initial position. A final minimization step without solute restriction was also carried out. Snapshots were collected every 20μs. Post processing of the trajectories (torsion, distances, RMSD) for further analysis was performed using the PTRAJ module of AMBER10[14].

For the polar/non-polar surface generations there have used the major conformers of the synthetic oligosaccharides in the presence of the mAbs using MAESTRO module of Schrödinger[16].

Sample preparation

The NMR study of the SF3a oligosaccharides in water solution was performed using 1mM samples in D_2O. Standard 2D-methods, such as TOCSY (with mixing times of 30 and 80ms), ROESY (300 ms), NOESY (400 and 600 ms), and HSQC experiments were employed, using Bruker Avance 500 and 600 MHz (equipped with a triple channel cryoprobe) NMR spectrometers.

Samples of the monoclonal antibodies IgG and IgM were lyophilized, and re-dissolved in deuterated phosphate-buffered saline (PBS, pH= 7.3) for the ligand binding experiments. The synthetic oligosaccharides and specific antibodies IgG C7-37 and IgM G19-2 were prepared at Institute Pasteur (Paris) and the synthesis has been reported elsewhere.

NMR studies (STD and trNOESY experiments)

Saturation transfer difference (STD)[12] experiments were performed on both 500 MHz and 600 MHz NMR spectrometers, at 298K or 313K. The experiments were carried out using samples with an antibody/ligand molar ratio of 1/200. In particular, the employed concentrations were 10uM/2mM for experiments with IgG, and 5uM/1mM for those with IgM). Different experiments (8000 scans each) were performed in which the saturation time of the protein was varied from 0.5 to 2s. Selective "on-resonance" irradiation of antibody resonances was performed at either δ7 or -1.5 ppm, while the "off-resonance" frequency was always set at 100 ppm. Similar results were obtained. STD experiments on free SF3a ligands at 1mM in D_2O were also acquired in order to obtain "blank" experiments.

All ligands displayed similar STD growth for all signals upon increasing of the saturation time. Therefore, the obtained results after 2 seconds of saturation time are discussed, since they displayed the best signal/noise ratio.

trNOESY experiments[17] were carried out on the 600MHz spectrometer, equipped with a triple-channel cryoprobe. The employed samples used a 1/30 antibody/ligand molar ratio, with a 33uM/1mM ratio for IgG and 20uM/600uM for IgM. The mixing time was 200 ms. trNOESY experiments were also performed to detect possible spin diffusion transfers. The obtained results were compared to those ROESY experiments recorded for the free ligand. Then, for both free (ROESY) and bound (trNOESY) states, the cross peak volumes were used to estimate the key interproton distances, using the isolated spin approximation. In all cases, an intraresidual proton pair that contained one of the protons involved in the distance to be determined was used as internal reference.

2.7 References

1. Boutet, J.; Guerreiro, C.; Mulard, L. A., Efficient synthesis of six tri- to hexasaccharide fragments of Shigella flexneri serotypes 3a and/or X O-antigen, including a study on acceptors containing N-trichloroacetylglucosamine versus N-acetylglucosamine. *J Org Chem* **2009,** *74* (7), 2651-70.
2. Allison, G. E.; Verma, N. K., Serotype-converting bacteriophages and O-antigen modification in Shigella flexneri. *Trends Microbiol* **2000,** *8* (1), 17-23.
3. Bryant, C. E.; Spring, D. R.; Gangloff, M.; Gay, N. J., The molecular basis of the host response to lipopolysaccharide. *Nat Rev Microbiol* **2010,** *8* (1), 8-14.
4. Raetz, C. R.; Whitfield, C., Lipopolysaccharide endotoxins. *Annu Rev Biochem* **2002,** *71*, 635-700.
5. Perepelov, A. V.; Shekht, M. E.; Liu, B.; Shevelev, S. D.; Ledov, V. A.; Senchenkova, S. N.; L'Vov V, L.; Shashkov, A. S.; Feng, L.; Aparin, P. G.; Wang, L.; Knirel, Y. A., Shigella flexneri O-antigens revisited: final elucidation of the O-acetylation profiles and a survey of the O-antigen structure diversity. *FEMS Immunol Med Microbiol* **2012,** *66* (2), 201-10.
6. (a) Burns, S. M.; Hull, S. I., Comparison of loss of serum resistance by defined lipopolysaccharide mutants and an acapsular mutant of uropathogenic Escherichia coli O75:K5. *Infect Immun* **1998,** *66* (9), 4244-53; (b) Joiner, K. A., Complement evasion by bacteria and parasites. *Annu Rev Microbiol* **1988,** *42*, 201-30.
7. Perepelov, A. V.; L'Vov V, L.; Liu, B.; Senchenkova, S. N.; Shekht, M. E.; Shashkov, A. S.; Feng, L.; Aparin, P. G.; Wang, L.; Knirel, Y. A., A similarity in the O-acetylation pattern of the O-antigens of Shigellaflexneri types 1a, 1b, and 2a. *Carbohydr Res* **2009,** *344* (5), 687-92.
8. Theillet, F. X.; Simenel, C.; Guerreiro, C.; Phalipon, A.; Mulard, L. A.; Delepierre, M., Effects of backbone substitutions on the conformational behavior of Shigella flexneri O-antigens: implications for vaccine strategy. *Glycobiology* **2011,** *21* (1), 109-21.
9. Theillet, F. X.; Frank, M.; Vulliez-Le Normand, B.; Simenel, C.; Hoos, S.; Chaffotte, A.; Belot, F.; Guerreiro, C.; Nato, F.; Phalipon, A.; Mulard, L. A.; Delepierre, M., Dynamic aspects of antibody:oligosaccharide complexes characterized by molecular dynamics simulations and saturation transfer difference nuclear magnetic resonance. *Glycobiology* **2011,** *21* (12), 1570-9.
10. Carlin, N. I.; Wehler, T.; Lindberg, A. A., Shigella flexneri O-antigen epitopes: chemical and immunochemical analyses reveal that epitopes of type III and group 6 antigens are identical. *Infect Immun* **1986,** *53* (1), 110-5.
11. Clement, M. J.; Fortune, A.; Phalipon, A.; Marcel-Peyre, V.; Simenel, C.; Imberty, A.; Delepierre, M.; Mulard, L. A., Toward a better understanding of the basis of the molecular mimicry of polysaccharide antigens by peptides: the example of Shigella flexneri 5a. *J Biol Chem* **2006,** *281* (4), 2317-32.
12. Mayer, M.; Meyer, B., Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew Chem Int Ed Engl* **1999,** *38* (12), 1784-1788.
13. Levine, M. M.; Kotloff, K. L.; Barry, E. M.; Pasetti, M. F.; Sztein, M. B., Clinical trials of Shigella vaccines: two steps forward and one step back on a long, hard road. *Nat Rev Microbiol* **2007,** *5* (7), 540-53.

14. Case, D. A.; Darden, T. A.; Cheatham, I. T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B. P.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A., AMBER 10. *University of California, San Francisco* **2008**.

15. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem* **2008,** *29* (4), 622-55.

16. Maestro, A Powerful, All-Purpose Molecular Modeling Environment, Version 8.5. *Schroedinger, LLC* **2008,** *New York*.

17. (a) Jimenez-Barbero, J.; Peters, T., NMR spectroscopy of glycoconjugates. *Wiley-VCH* **2003**; (b) Claridge, T. D. W., High-Resolution NMR Techniques in Organic Chemistry. *Tetrahedron Organic Chemistry Series* **1999,** *19*; (c) Roldós, V.; Cañada, F. J.; Jimenez-Barbero, J., Carbohydrate-Protein Interactions; A 3D view by NMR. *Chembiochem* **2011,** *12*, 1-17.

***Chapter 3*.** Molecular Recognition of Complex-Type Biantennary *N*-Glycans by Protein Receptors: a Three-Dimensional view on Epitope Selection by using NMR

3.1 Introduction

Molecular recognition is at the heart of essential biological events. The recognition of saccharides presented as part of natural scaffolds (protein, sphingolipids) by protein receptors (lectins) mediates key processes underlying many aspects of life as reflected by the term "sugar code"[1]. Matching the wide spectrum of glycan structures (glycome complexity[2]), lectins are found in all branches of the phylogenetic tree and have proven to be versatile tools for glycan detection and structural characterization[3].

Indeed, there is nowadays particular interest in the identification of glycoprotein glycan structures as a result of altered glycoprotein expression occurring in diseases, hereby defining functional glycobiomarkers of diagnostic value[4]. Recent technological advances have led to the development of lectin and glycan microarrrays, which offer a comparative view of the lectin–glycan interactions and provide a high-throughput approach to screen a wide range of glycan profiles and to determine glycan binding specificity of the lectin[5].

With the growing realization that the presentation of a glycan epitope can affect its recognition by lectins[3b], it is timely to study these interactions with natural glycans of the proper size and then rigorously define the contact sites, especially in cases with more than one possible interaction mode.

NMR and X-ray crystallography provide key information at atomic resolution on biomolecules, including complexes. However, in the protein–carbohydrate interaction field, these techniques have mainly focused on acquiring information of complexes formed by lectins with small/medium-sized fragments of natural glycan chains. Often, it is an open question how affinity/specificity is regulated on the level of complex glycans.

3.2 Objectives

N-Glycans[6] are common modifications of membrane/secreted proteins. They confer specific properties to the associated protein that will ultimately determine its fate. The structures of *N*-glycans are widely diverse, although share common structural motifs. They always contain a core pentasaccharide unit, which is composed of a

chitobiose unit linked to a trimannoside moiety that introducesa branching point with two antennae. Both antennae can be extended by *N*-acetyllactosamine (LacNAc) units, terminated by α2,6- or α2,3-linked sialic acid. As abundant species among *N*-glycans, we have herein studied the biantennary complex-type nonasaccharide–Asn (with LacNAc termini) (Figure 3.1 compound 2) and its α2,6-sialylated undecasaccharide–Asn derivative (Figure 3.1 compound 4), which were isolated from egg yolk by the group of Prof. C. Unverzagt in Bayreuth (Figure 3.1). These structures represent the common motifs of *N*-glycans on glycoproteins, with α2,6-sialylation acting as the salient signal for masking/docking[1b, 7]. In fact, these structures contain several binding epitopes in their primary structures and therefore are very good tools to understand how lectins select a given epitope among all those present in the carbohydrate ligand. From the receptor side, different lectins have been selected. According to published data, they should be able to bind different epitopes of the natural *N*-glycans. In particular, two agglutinins described with binding affinity to the *N*,*N*′-diacetylchitobiose ($(GlcNAc)_2$) part of the core have been chosen, namely, wheat germ agglutinin (WGA)[8] and the WGA-derived single domain similar to hevein[9]. Hevein, one of the smallest plant lectins with 43 amino acids, was the first lectin whose structure in complex with its sugar ligand was solved in solution[10] by using NMR techniques[8]. As mentioned in the introduction, WGA harbors eight hevein domains and displays a dual specificity for both GlcNAc and Neu5Ac,[8] which are monosaccharides found at opposing ends of natural *N*-glycans. Whether both sugars are binding targets in an *N*-glycan or whether the lectin exhibits selectivity was not known, prompting us to perform this study.

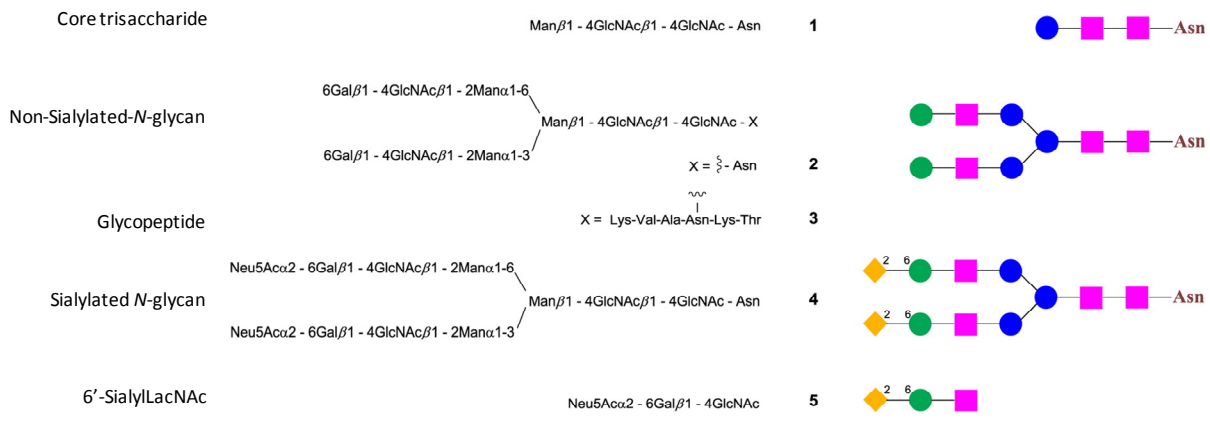


Figure 3.1 Glycan structures used in this study.

The used approach can likewise be used to obtain structural view on target–site interaction. To illustrate this application, the *Viscum album* agglutinin (viscumin, VAA) was selected. It binds LacNAc termini on *N*-glycans, and also in the case of α2,6-sialylated *N*-glycans[11]. As a negative control, a mixture of lectins, for whom α2,6-sialylgalactose terminus blocks recognition, was chosen. In this context, the two agglutinins from the seeds of *Maackia amurensis* (the leukoagglutinin MAL and the hemeagglutinin MAH) have been selected[8, 12].

We show that the mode of epitope presentation is crucial for glycan recognition and that the selectivity observed in assays with free sugars can change after subtle alterations of the *N*-glycan structure. The same can hold true for lectins differing in quaternary structure.

3.3 Results

3.3.1 The single hevein Domain vs WGA

WGA constitutes the classical model for chitooligosaccharide binding. Additionally, WGA is often included among the sialic acid binding lectins. Therefore, two potential epitopes are presented in sialylated glycopeptides for this lectin. In fact, the recognition features of chitooligosaccharides by hevein domains have been well established through NMR[9-10]. More recently, we have demonstrated that the trisaccharide *N*-glycan core (Figure 3.1 compound 1) is also recognized by hevein, with GlcNAc1 (linked to Asn) on subsite +2 (on top of W21), GlcNAc2 on subsite +1 (on top of W23), and βMan at subsite −1, leaving the Asn residue outside of the binding site, without establishing interactions with the lectin[13]. This structural model was used as a template for analyzing the recognition of complex-type *N*-glycans. However, when we titrated a 0.3 mM solution of a single hevein domain (the B domain of WGA[14]) with increasing amounts of the sialylated *N*-glycan–Asn (Figure 3.1 compound 4), no changes were observed for the ^{1}H NMR resonances of the lectin. Similarly, attempts to dock the undecasaccharide–Asn (Figure 3.1 compound 4) using the binding mode of the core trisaccharide–Asn (Figure 3.1 compound 1 and Figure 3.2) wereunsuccessful. In fact, branching at position 6 of the central βMan residue causes a steric clash of the

sugar extension with different protein residues, independent of the presence of either *gg* or *gt* rotamers at the C5–C6 linkage. In contrast, WGA is known to bind *N*-glycans[8].

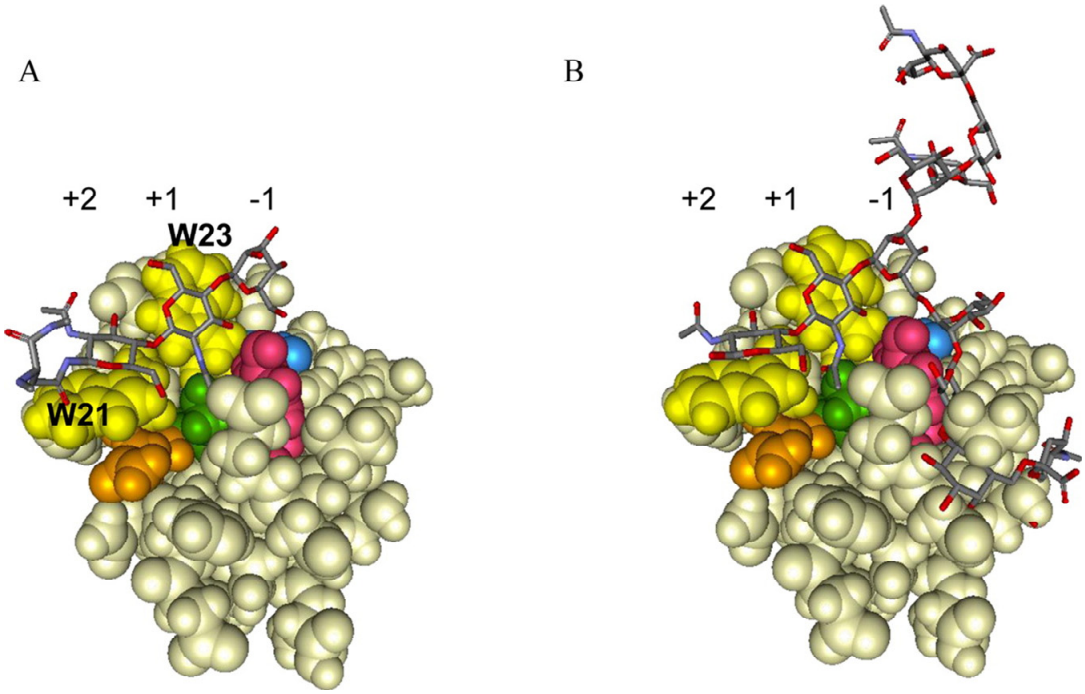


Figure 3.2 (A) Complex of hevein with the trisaccharide core (Figure 3.1 compound 1) as deduced from NMR[13]. (B) Attempted docking of the extended *N*-glycan structure (Figure 3.1 compound 4) obtained by superimposition of the sialylated *N*-glycan in the trisaccharide core template (see panel A). The extension at position O6 of the central Man moiety in the core causes steric clashes with the protein. The *gg* orientation is shown at Man C5–C6. The *gt* rotamer is sterically also forbidden. Subsites are indicated on top, and the key amino acids, i.e., Trp21 and Trp23, are annotated.

WGA is a dimeric lectin where each subunit consists of an assembly of four hevein domains[15]. It has been described to be specific for terminal Neu5Ac and GlcNAc moieties by inhibition studies[8], X-ray crystallography[15], and NMR spectroscopy[16]. The structure of complexes, also with a small sialoglycopeptide[15c], reveals that the primary binding site is constituted by three conserved aromatic residues from one hevein domain. Nevertheless, further polar residues (S114 and E115) from the neighboring hevein domain contribute to the stabilization of the complex, providing structural evidence for the intra-protein cooperation of lectin domains.

Strikingly, and in contrast to the observations for the single hevein domain, the STD experiments performed with the sialylated *N*-glycan–Asn (Figure 3.1) in the presence of the multidomain lectin showed STD signals, evidencing the existence of interaction. Analysis of the STD spectra (Figure 3.3) clearly showed that the terminal

Neu5Ac residue was not the key point for the recognition process. Only a very weak effect was observed for the methyl group of the 5-acetamide, while the ring hydrogens were not present in the STD spectrum (Figures 3.3 and 3.4).

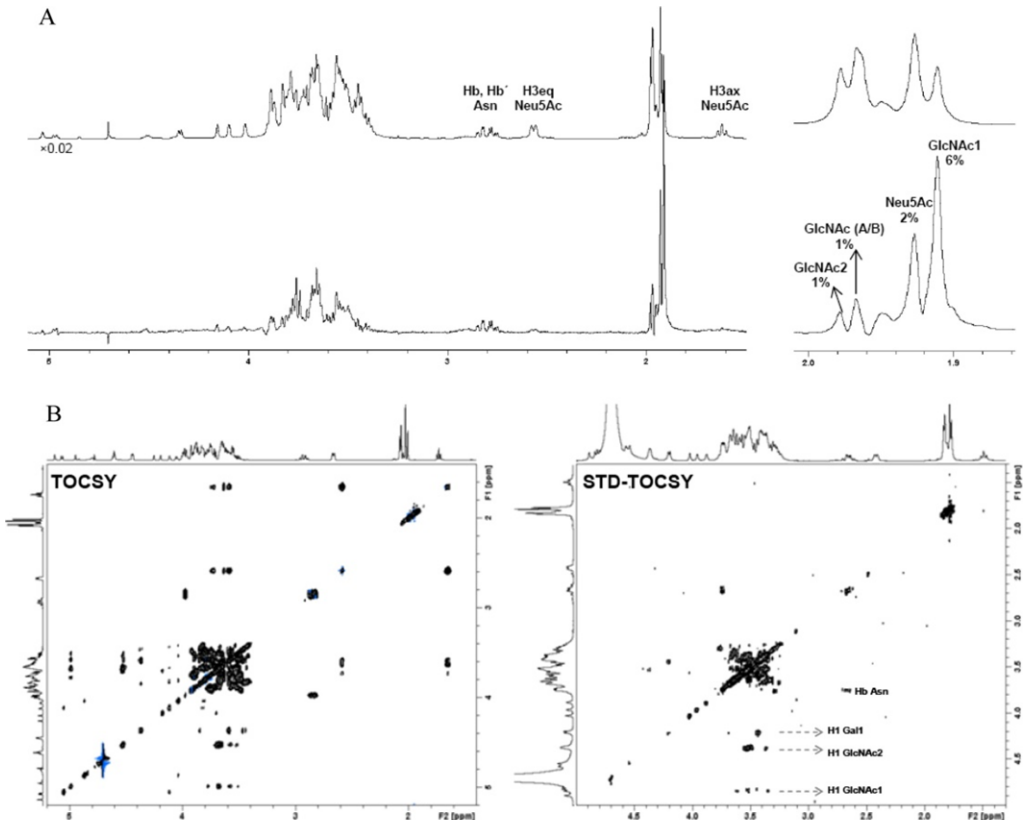


Figure 3.3 (A) STD (below) and off-resonance spectrum (on top) of the sample containing WGA and the sialylated *N*-glycan (Figure 1**)**. The region of the methyl protons of the Ac groups is enlarged on the right. (B) left, TOCSY of ligand 4; right, STD-TOCSY of the mixture of WGA and sialylated *N*-glycan 4. The sialyl residue does not contribute to the binding event. In contrast, the Asn residue provided clear STD signals indicating its direct involvement in the recognition process.

In fact, the residues in closer contact with the protein are at the opposite end of the molecule, namely, the GlcNAc1 and Asn moieties. Thus, the binding mode is different from that occurring between a single hevein domain (Figure 3.2A) and the trisaccharide core (Figure 3.1), in which the Asn moiety was far from the lectin (positive NOEs in the NOESY spectrum of the complex)[13]. For hevein itself, GlcNAc1 and GlcNAc2 are located at subsites +2 and +1, respectively. However, for WGA, the docking/minimization procedure based on the experimental observations showed that the binding mode is shifted by one subsite: the Asn residue now occupies subsite +2 (on top of W21); GlcNAc1 is on subsite +1 (on top of W23), and GlcNAc2 sits on subsite −1. In this situation, the βMan residue is shifted one subsite away from the center of

binding to the protein, and therefore, the typical branching at O6 does not cause steric clashes with the protein (Figure 3.5). Obviously, the following question arises: why does the single hevein domain not use this binding mode for recognizing the large sialylated*N*-glycan? Detailed comparisons pointed to the presence of the additional polar residues in the neighboring hevein domain in WGA; that is, S114 and E115 appear as key factors stabilizing the complex. They form hydrogen bonds with the *N*-glycan (Figure 3.5), which are absent in the single hevein domain. Probably, this binding mode (GlcNAc1–Asn on subsites +1 and +2, respectively) is energetically not favorable in the case of just a single hevein domain[13], where these new interactions cannot contribute to the binding enthalpy.

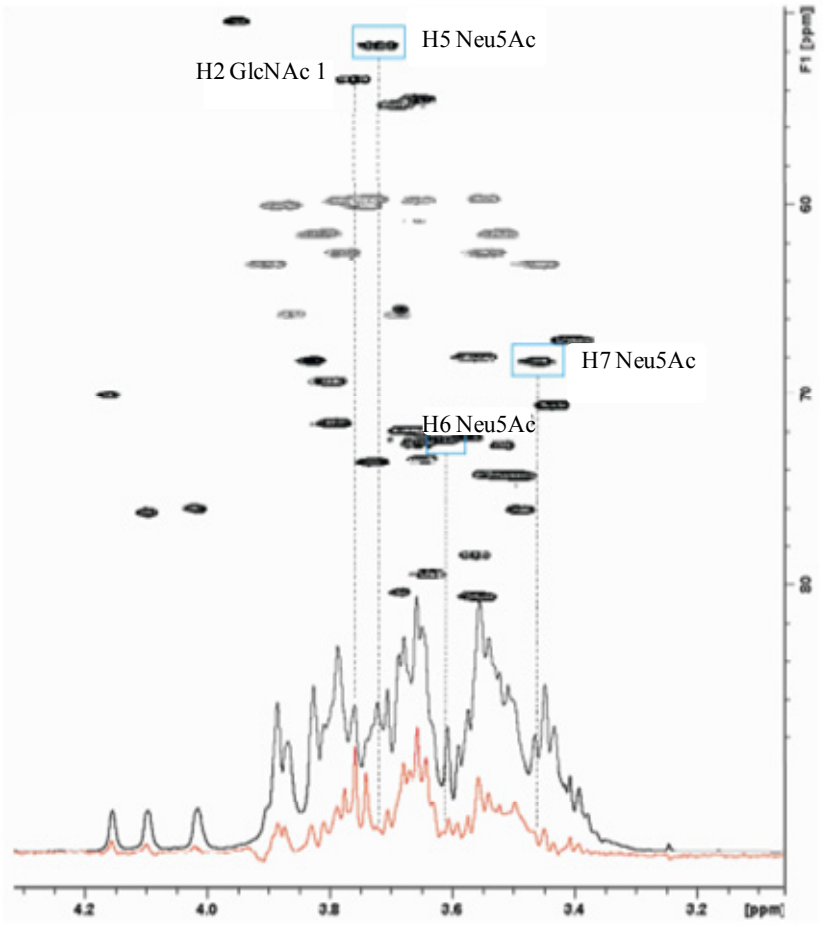


Figure 3.4 HSQC (edited) spectrum of compound in Fig. 3.1, and superimposed STD (in red) and off resonance spectrum (in black) in presence of WGA. Hydrogens of the sialyl residue do not appear in the STD, implying that they are not directly involved in WGA binding.

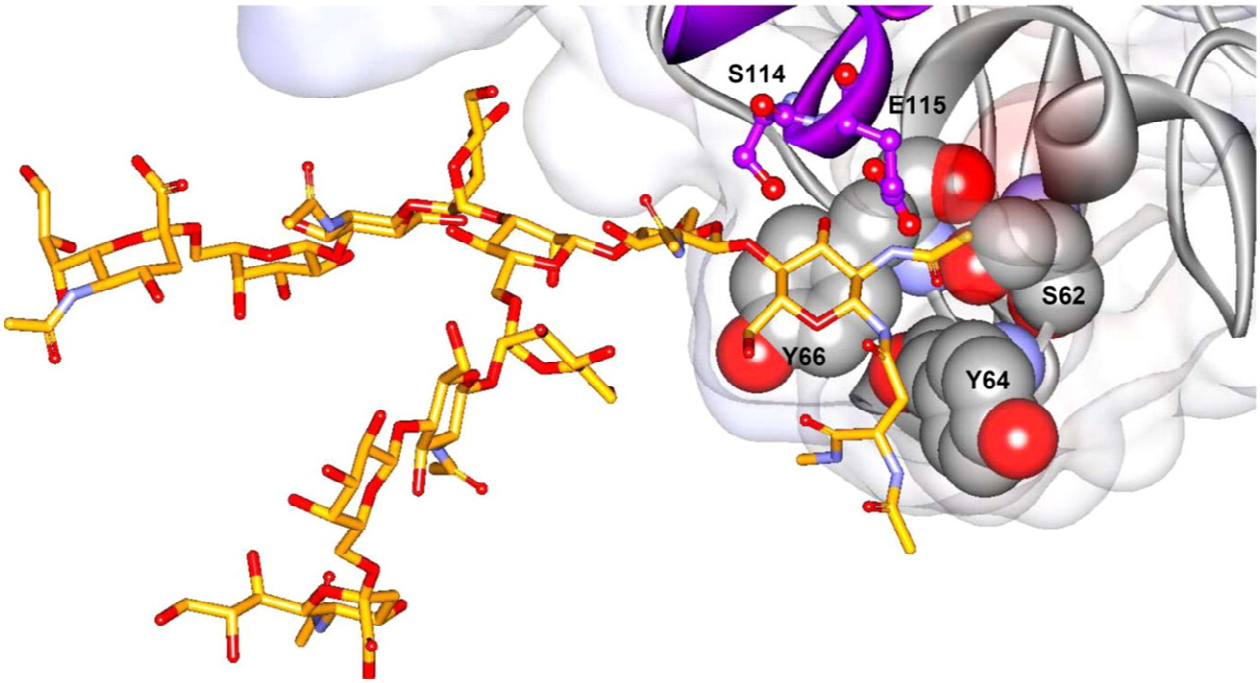


Figure 3.5 Deduced complex of the sialylated *N*-glycan (in orange) in the primary binding site of WGA, as deduced from the docking, minimization, and MD process. Amino acids Y66, Y64, and S62 (equivalents to W21, W23, and S19 in hevein) from one hevein domain are represented in gray CPK, while S114 and E115 from the neighboring hevein domain are represented in balls and sticks, in purple. The sialyl residues are exposed to the solvent and not directly involved in WGA binding.

Analogous STD experiments were carried out with the non-sialylated *N*-glycan–Asn (Figure 3.1 compound 2 and Figure 3.6). They were employedas control to ascertain the binding mode deduced above. The results were very similar to those obtained with the sialylated *N*-glycan: Asn and GlcNAc1 displayed the strongest STD intensities, with much lower intensities for the branches.

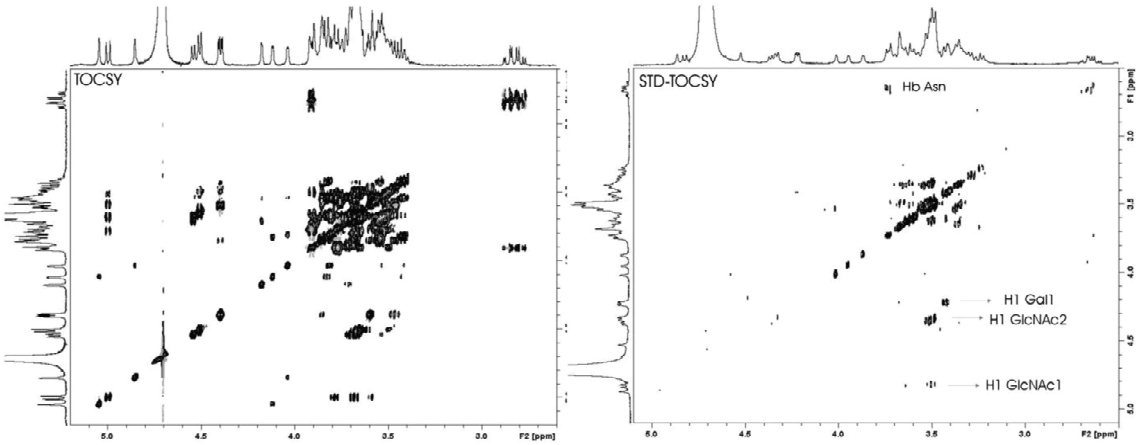


Figure 3.6 (left) TOCSY spectrum of non-sialylated *N*-glycan. (right) STD-TOCSY of WGA/non-sialylated *N*-glycan. As for sialylated *N*-glycan, the Asn and GlcNAc1 residues provided clear STD signals indicating their direct involvement in the recognition process.

As already mentioned, WGA is often included among the sialic acid binding lectins. However, our experimental STD data have shown that, although sialylated *N*-glycan is indeed recognized by WGA, the terminal sialic acid of the undecasaccharide does not establish primary interactions with the lectin. Instead, the interactions occur via the GlcNAc1 and Asn moieties at the stem region of the glycan.

To prove binding to a sialylated, but GlcNAc-free compound, one additional STD experiment with 6′-sialylLacNAc (Figure 3.1 compound 5) was acquired. The STD spectrum (Figure 3.7) clearly showed that only the protons from the sialyl (Neu5Ac) residue showed STD intensity, especially those from the acetyl group, in agreement with the published X-ray crystallographic data[15]. The other STD signals corresponded to H5, H6, H7, and H4 of Neu5Ac. Thus, although WGA recognizes simple sialyl-containing oligosaccharides, when presented in the context of a complex *N*-glycan, the sialic acid is therefore not its primary contact site, as it is for the isolated branch-end trisaccharide. Explicitly, in the sialylated *N*-glycan–WGA complex, only the acetyl group of Neu5Ac showed weak STD, and protons H5, H6, and H7 did not show STD signals (See Figure 3.4).

This result underscores the inherent value of the given approach to team up binding-site identification by STD NMR spectroscopy with *N*-glycan preparation. When more than one possible contact site is presented, the given strategy can determine the preferred epitope in the context of the *N*-glycan.

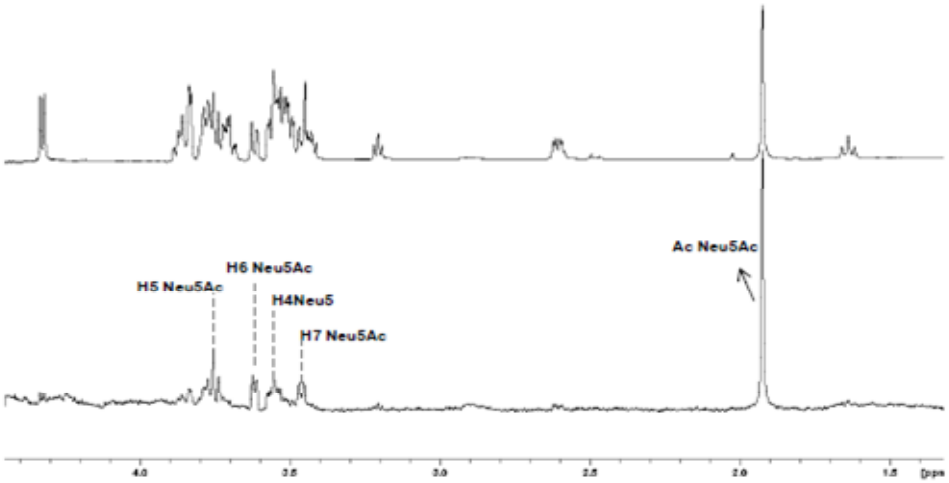


Figure 3.7 STD spectrum (below) of 6'-sialylLacNac with WGA. On top, off resonance reference spectrum. Hydrogens of the sialyl residue clearly appear in the STD, implying the direct involvement in WGA binding.

Since the recognition of the sialylated and non-sialylated-*N*-glycan by WGA mainly involves GlcNAc1 and the Asn residue, it remained to be clarified whether there is an influence of a peptidic moiety, as occurring in glycoproteins. Thus, an STD experiment was carried out under the same experimental conditions for a mixture of the glycopeptide (Figure 3.1 compound 3) and WGA. The spectrum showed STD intensities similar to those obtained for the non-sialylated-*N*-glycan (Figure 3.8). In particular, the protons of the additional peptide residues did not show up in the spectrum, indicating exclusive recognition of the glycopeptide by WGA through the GlcNAc–Asn region. We next proceeded to a case study of actual branch-end recognition.

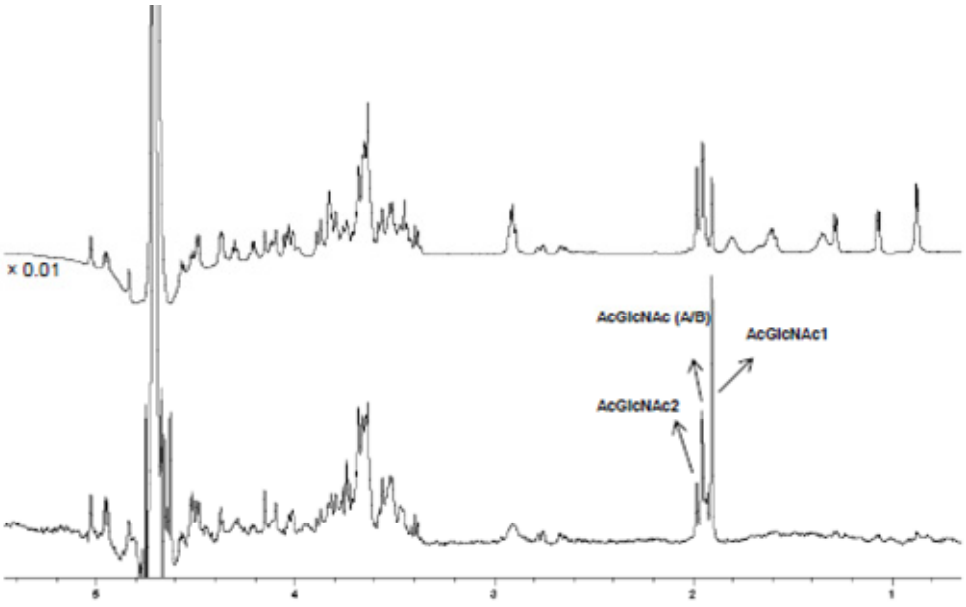


Figure 3.8 STD (below) and off-resonance (on top) of glycopeptides 1 and 3 and WGA. As for 2 and 4, the Asn and GlcNAc1 residues provided clear STD signals indicating their direct involvement in the recognition process.

3.3.2 Viscumin

As mentioned in the introduction section, this lectin is a toxic AB-type protein. The B chain harbors two carbohydrate-binding sites, characterized by W38 and Y249. In solution, the lectin activity of VAA can be attributed primarily to the Tyr site[17]. The recognition mode of lactose by VAA has been characterized by X-ray crystallography, chemical mapping, and modeling[18]. It has been reported that VAA strongly binds Neu5Acα2–6Galβ1–4GlcNAc on gangliosides and glycoproteins, while gangliosides

with terminal galactoses are poorly recognized[19]. α2,6-Sialylation can improve inhibitory capacity of lactose[11b]. Nevertheless, a structural rationalization to those facts has not yet been given.

Figure 3.9 shows the STD spectra, along with their corresponding reference spectra, for VAA in the presence of the sialylated *N*-glycan (on top) and of the non-sialylated analogue (below). Both compounds are bound by VAA, but the STD intensities were significantly stronger for the sialylated than for the non-sialylated. Since strictly the same experimental conditions were kept, the results suggest respective grading of interaction for VAA. Furthermore, the distinct protons of the Neu5Ac residue (H3ax, H3eq, and Ac) clearly revealed that this residue is in close contact with VAA. This fact was confirmed by the STD-TOCSY spectrum (Figure 3.10). Thus, in this case, the terminal Neu5Ac moiety is clearly involved in the recognition process, and its absence weakens but does not prevent glycan–ligand interaction.

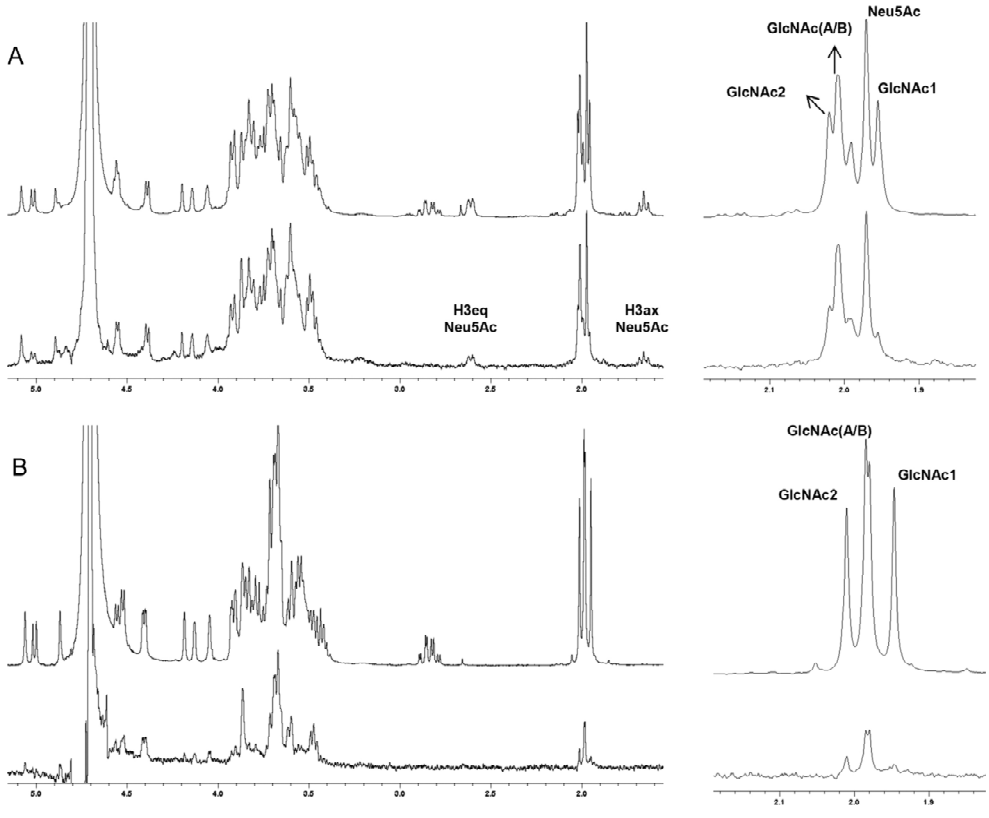


Figure 3.9 STD spectra of the mixtures of VAA with (A) 4 and (B) 2. The methyl groups regions are enlarged at the right-hand side, and the resonances are annotated for each case. The STD spectra are shown below, while the corresponding off-resonance spectra are on top. In all cases, VAA binds both molecules.

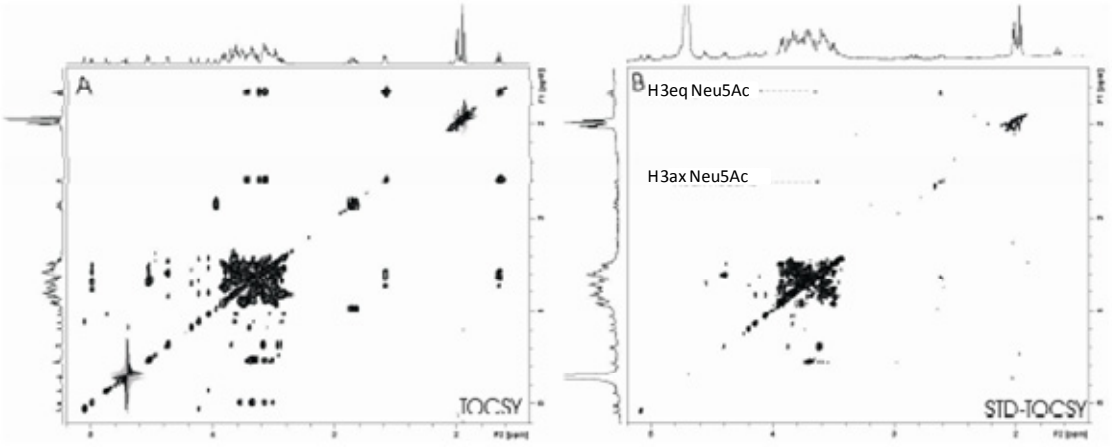


Figure 3.10 (A) TOCSY spectrum of sialylated *N*-glycan. (B) STD-TOCSY with Viscumin. The sialyl residue is in contact to the lectin surface.

Three-dimensional (3D) structures of the corresponding complexes were generated by using molecular modeling techniques, taking the available experimental X-ray coordinates for the VAA–lactose complex as a starting geometry. In the resulting 3D structure, the galactose-binding site is formed by amino acids D335, N256, and Q238 that establish hydrogen bonding with the Gal residue and by Y249, which establishes stacking interaction with the α-face of the Gal moiety, explicitly with H3, H4, and H5 (Figure 3.11A). Inspection of this experimental geometry detects space for a sugar extension at Gal O6. This location is surrounded by polar amino acids that may efficiently interact with sialic acid residues; they are S200, R245, and Q238. Docking of sialylated *N*-glycan at this locus, using the lactose-binding model as a template, provided a stable pose, which was further optimized through an energy minimization and MD protocol. Of note, further interactions can then take place between Neu5Ac and the above-mentioned polar residues (Figure 3.11B): R245 establishes a bidentate hydrogen bond with the glycerol chain of the sialic acid residue, S200 is hydrogen bonded to the carboxylic acid, and Q238 establishes a cooperative hydrogen bond with OH7 of Neu5Ac and OH3 of the branching GlcNAc. Evidently, these interactions appear to underlie the preferential affinity reported above. Terminal LacNAc units are recognized, but further sialylation to Neu5Acα2–6LacNAc enhances binding.

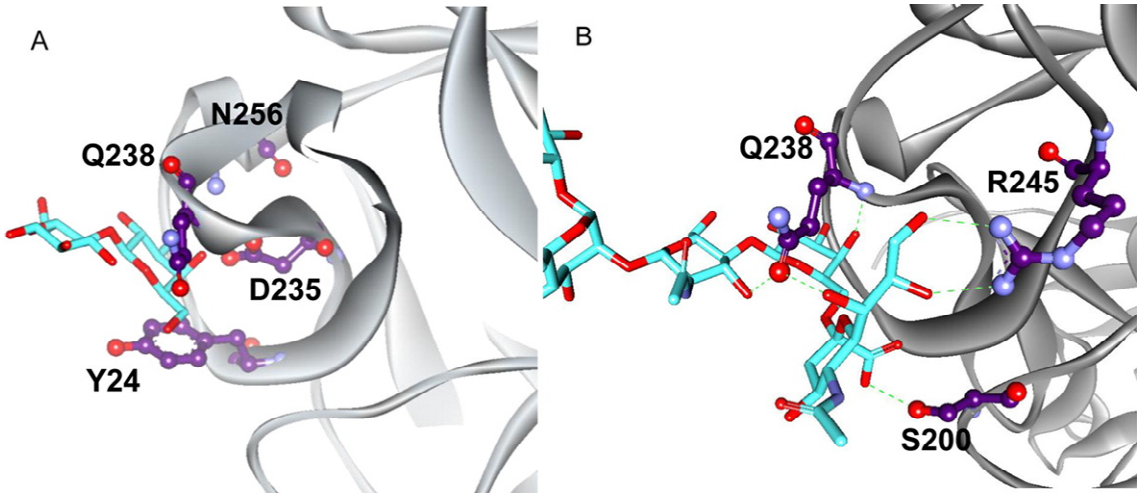


Figure 3.11 (A) X-ray structure (pdb code 1PUU) of VAA complexed with lactose. The amino acids implicated in interactions with the ligand are highlighted. (B) Docking pose of the sialylated *N*-glycan on the Tyr site in subdomain 2γ. The amino acids providing new further interactions (hydrogen bonds depicted in green) with the Neu5Ac residues are highlighted.

3.3.3 *Maackia amurensis* lectins

The mixture of the two lectin species from the seeds of *Maackia amurensis*, MAL and MAH, was employed[12]. Both isolectins are known to be able to bind to α2,3-sialylgalactose, although with different preferences: the leukoagglutinin (MAL) is known to preferentially react with Neu5Acα2,3Galβ1,4βGlcNAc moieties present in *N*-glycans[16c], while the hemagglutinin (MAH) has higher affinity for the disialylated tetrasaccharide Neu5Acα2,3Galβ1,3[Neu5Ac(2,6)]αGalNAc present in *O*-glycans[4e, 12, 20]. The binding sites of both isolectins are very similar, with the only differences being the amino acids Y221 and E222 of MAL, which are replaced by Ala residues in MAH. The X-ray crystal structure of MAL complexed with 3′-sialyllactose has been reported[21].

1D STD and two-dimensional STD-TOCSY experiments were acquired for the sialylated and non-Sialylated-*N*-glycan, using samples with a protein/ligand 1:20 (50 μM/1 mM) ratio. The corresponding spectra are shown in Figure 3.12. The comparison of the spectra on the left (sialylated *N*-glycan-Asn) and on the right (non-sialylated) enabled the assessment that the MAL–MAH mixture indeed does not recognize the Neu5Acα(2,6)Gal-containing molecule (Figure 3.1 compound 4), providing a negative

control for specificity, while it does efficiently bind to the non-sialylated species. The documented lack of reactivity in lectin-affinity chromatography or in glycan arrays serves as further control[4e, 12].

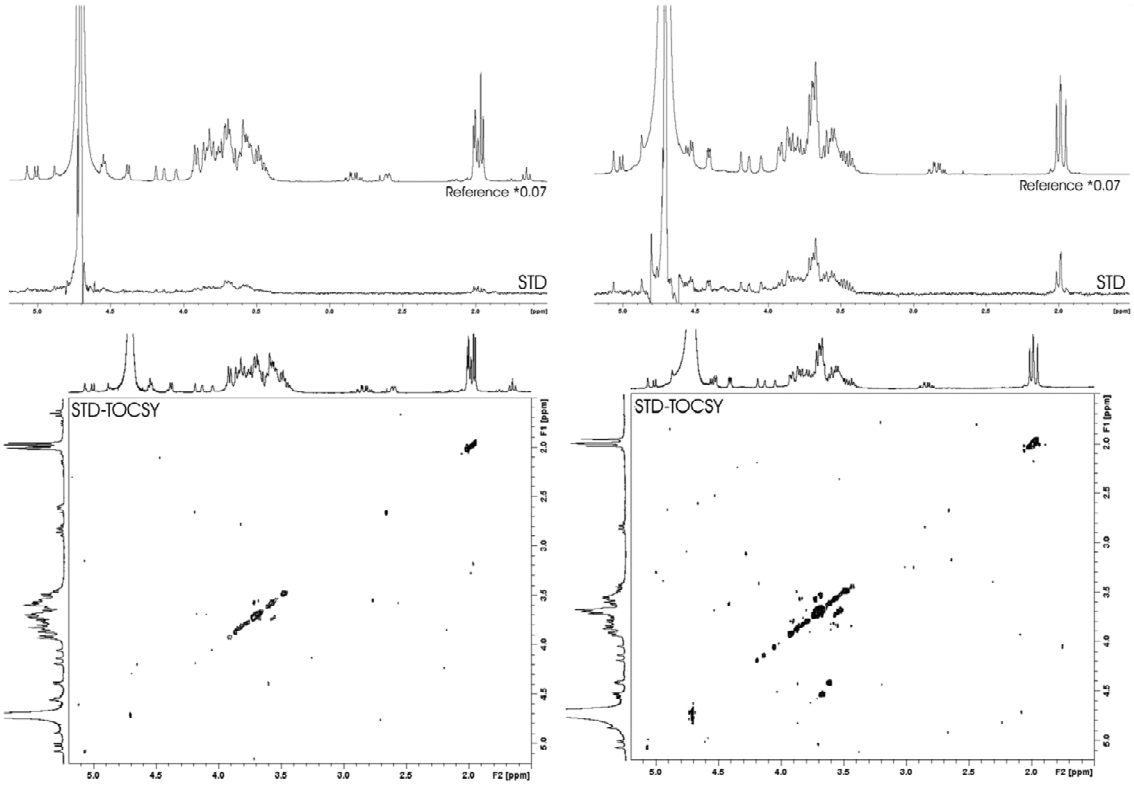


Figure 3.12 (left) STD and STD-TOCSY spectra corresponding to MAA–sialylated *N*-glycan (Compound 4). No STD is detected. (right) Spectra corresponding to MAA–non-sialylated *N*-glycan (Compound 2). The terminal *N*-acetyllactosamine residues provided clear STD signals indicating their direct involvement in the recognition process.

From the analysis of this STD spectrum, it was deduced that the residues interacting with the lectin are the non-reducing end terminal galactose and GlcNAc residues (LacNAc unit), while in this case, the GlcNAc1 and Asn stem at the other end is not in contact with the protein. This observation is in agreement with the previously reported specificity data and with the published crystallographic structure (pdb code 1DBN). Indeed, the binding site of MAL is unusually narrow for a lectin. Both the terminal galactose and GlcNAc residues are sandwiched between Y136 and Y221, leaving Gal O6 deep inside the binding site and establishing hydrogen bonding with D137, clearly precluding the extension through Gal O6, as it is the case of the 2,6-sialylated *N*-glycan. Now, the absence of the sialic acid moiety allows the terminal LacNAc unit of non-sialylated-*N*-glycan to perfectly fit inside the binding pocket,

following the typical feature for lectin binding[3b] by establishing a combination of CH–π stacking interactions with Y136 and Y221 and hydrogen bonds with Y131, D87, D137, and E224[21]. Further interactions are established with GlcNAc2 and ManB through Y136 and Y221 (Figure 3.13A). Terminal elongation of the non-sialylated-*N*-glycan with Neu5Ac but with an α2–3 linkage was then explored by modeling procedures (Figure 3.13B), yielding a complex in which Neu5Ac not only fits in the narrow lectin binding site but further contributes to the binding by hydrogen bonding to amino acids K107, S104, and Y136, as also observed in the X-ray structure for sialylα2,3-lactose[21]. Thus, in this particular example, the presence of the α2,6-linked Neu5Ac moiety impairs the *N*-glycan ligand binding to the lectin. Interestingly, in the absence of Neu5Ac, recognition of the underlying LacNAc unit occurs.

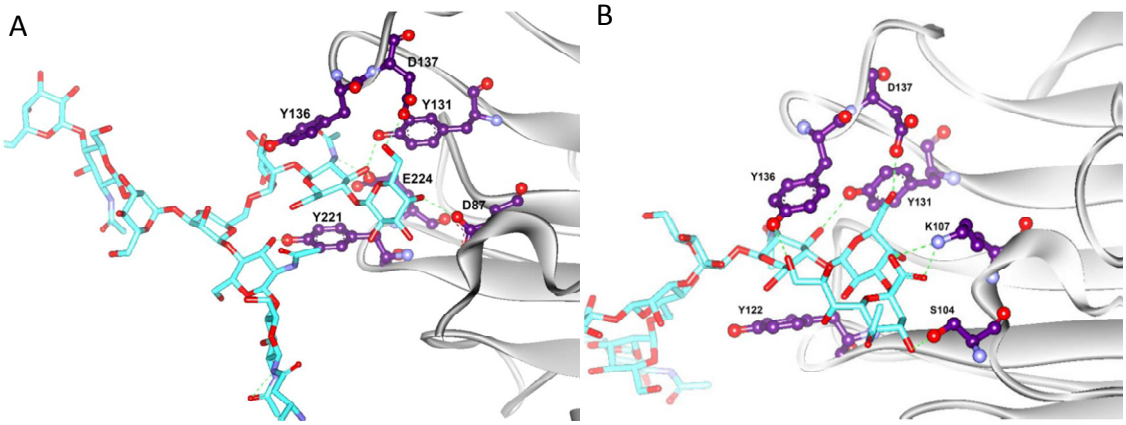


Figure 3.13 (A) Modeled structure (using a docking and MD protocol) for the complex formed between MAL and the non-sialylated-*N*-glycan, using the deposited X-ray structure (pdb code 1DBN) as the template structure. The *N*-acetyllactosamine at the 1→6 branch has been docked at the binding site. (B) 3D model for the recognition of an analogous undecasaccharide as the sialylated-*N*-glycan but with terminal Neu5Ac in α2-3 linkages instead of the trisaccharide Neu5Acα2-3Galβ1-3GlcNAcβ with the 1DBN X Ray structure and further energy minimization.

3.4 Conclusions

The preparation of *N*-glycans facilitates definition of the docking sites of lectins on complex glycans. Shown herein, the nature of the target epitope strictly depends on the nature of the tested lectin, which can select distinct determinants of the complex-type saccharide chain with exquisite specificity, depending on the relative presentation of the different residues and the architecture of the binding site.

We have underscored the importance of the coexistence of binding sites from different lectin domains for achieving glycan binding by comparing a single hevein domain and WGA. The monomeric unit does not recognize the large glycan, with the branching at the βMan residue being the key structural element for precluding the interaction. In contrast, the multidomain lectin WGA binds the terminal reducing end of either non- or sialylated-*N*-glycan, namely, the GlcNAc–Asn epitope, owing to the possibility of achieving additional inter-site stabilizing interactions from one of the neighboring hevein domains. In this context, it is worth emphasizing that WGA does not bind to non- or sialylated-*N*-glycan through the $(GlcNAc)_2$ moiety, as it would be expected. The chemical nature of the branched *N*-glycans excludes the possibility of efficient interactions of the chitobiose stem with the two consecutive aromatic residues typical for hevein domains, and thus, the recognized epitope is shifted by one position to the GlcNAc–Asn fragment. Despite being considered as a sialic acid binding lectin, our data have clearly shown that the sialyl residue of the sialylated-*N*-glycan is not a key player in the binding process. Thus, in naturally occurring *N*-glycopeptides, although presented at the spatially less readily accessible non-reducing end, binding occurs via the *N*-glycan–Asn stem region, a result of conspicuous importance for the interpretation of data using WGA in array platforms or in cyto- or histochemical glycophenotyping. Given the assumed significance of sialylation, for example, in tumor suppressor activities in response to environmental effectors or in immunity[22], caution should thus be exercised regarding WGA as a tool to detect alterations in this parameter. As demonstrated, on the level of small glycans, WGA can accommodate sialic acid residues. Indeed, using α2,6-sialyllactose, we clearly demonstrate that this small trisaccharide is recognized by WGA, but the situation is different on the level of *N*-glycans. Considering the importance of terminal sialylation, we have extended our study by analyzing the respective recognition structurally.

Work with VAA has yielded insights into how α2,6-sialylation can be accommodated by this lectin, providing enhanced binding in natural *N*-glycans when compared to the non-sialylated LacNAc-terminated analogue. Our methodology has permitted identification of the direct involvement of the sialic acid, as well as a network of interactions.

Conversely, the results for the MAL–MAH mixture have permitted definition of the role of the α2,6-linked sialic acid residue in blocking binding of the corresponding

sialylated-*N*-glycan. The binding to the LacNAc terminus of the non-sialylated-*N*-glycan is detectable, providing the control reference for positive binding.

In conclusion, the wide variety found both within the glycome and lectin architectures is a challenge for analyzing carbohydrate–lectin interactions. Advances in the macroscopic analysis of oligosaccharide binding using array-based and other technologies[1b] can strategically be complemented by the given approach, to rigorously define the contact sites also of complex glycans. Thus, the combination of *N*-glycan preparation, STD NMR spectroscopy, and modeling protocols gives new insights and proves its potency to detect and characterize epitope selection and interaction. They are of particular relevance for lectins used for biomedical applications.

3.5 Materials and Methods

Methods

NMR experiments

The ^{1}H NMR resonances of the ligands were completely assigned through standard TOCSY (60 ms mixing time), NOESY (300 and 500 ms mixing time), and HSQC experiments on 800 MHz (cryo) and 600 MHz spectrometers. Typical concentrations were 1 mM for the homonuclear experiments and 2 mM for the heteronuclear experiments.

The samples for saturation-transfer difference (STD) experiments as adapted to the individual lectins[23] were prepared in phosphate-buffered saline (pH = 5.7 for MAA and WGA, and pH = 7.3 for Viscumin) using ligand/lectin ratios varying from 1:20 to 1:100. The applied temperatures varied between 283 and 303 K. Molar ratio and temperature were optimized in each case. Representative experiments with significant STD responses are presented in the figures.

Data from one-dimensional (1D) STD experiments were acquired at 600 and 500 MHz using a Gaussian pulse (49 ms) cascade separated by 1 ms delays. STD-TOCSY experiments were acquired at 500 MHz, using 256 increments and a isotropic mixing time of 60 ms. In all cases, the on-resonance frequency was set at the aromatic or

aliphatic regions (6.9 or 0.5 ppm, respectively), and a series of experiments were recorded, where the saturation time was varied between 200 ms and 2 s. The off-resonance frequency was always set at 100 ppm. Under these conditions, the free ligand in solution showed residual STD intensities in the 1D STD spectra that were taken into account when analyzing the STD spectra.

Molecular Modeling

The templates for the modeling procedures were built based on X-ray structures of WGA complexed with small oligosaccharides (pdb codes: 2UVO and 2CWG), 1DBN for the *M. amurensis* isolectin and 1PUU for viscumin. The structure of the *N*-glycan was superimposed in the binding site working with the most populated conformation found for the free state (according to a standard NOE/molecular modeling approach). The complex structure was then submitted to a short molecular dynamics (MD) run, followed by energy minimization with a low gradient convergence threshold (0.02) in 5000 steps. In all cases, the OPL2005 force field[24] was employed, as integrated in the Schroedinger MAESTRO suite of programms[25].

Lectins

WGA and MAA were purchased from Sigma–Aldrich, VAA was obtained and controlled for purity as described[26], and WGA-B domain was also obtained as previously described[14].

Ligands

The 6'-SialylLacNAc was purchased from Sigma–Aldrich. The core trisaccharide[13], Non- and Sialylated-N-Glycan were obtained as previously described[27]. The Glycopeptide was prepared from its corresponding sialylated undecasaccharide–glycopeptide, obtained from egg yolk in analogy to published procedures[28]. This undecasaccharide–glycopeptide (5.7 mg, 2 μmol) was dissolved in 400 μL of phosphate buffer (75 mM; pH 6), and a solution of 0.2 mg (0.6 U) of neuraminidase from *Clostridium perfringens* (Sigma–Aldrich, EC: 3.2.1.18) in 50 μL of phosphate buffer (75 mm, pH 6) was added. The mixture was incubated at ambient temperature for 3 days (TLC: 2-propanol/1 M ammonium acetate 1.5:1) and cleared by centrifugation, and the supernatant was purified by gel filtration (Superdex 30 (1.6 cm × 60 cm); flow rate, 1 mL/min; eluent, 0.1 *m* NH_4HCO_3; detection, 214 nm). The fractions eluting at 77

min were collected, lyophilized, and desalted by gel filtration (Sephadex G25 (2.5 cm × 70 cm); flow rate, 1 mL/min; eluent, 5% ethanol in water; detection 214 nm). The fractions eluting at 201 min were lyophilized and yielded 4.05 mg of Non-Sialylated-*N*-Glycan glycopeptide (98.2%).

3.6 References

1. (a) Haltiwanger, R. S.; Lowe, J. B., Role of glycosylation in development. *Annu Rev Biochem* **2004,** *73*, 491-537; (b) Gabius, H., The sugar Code: Fundamentals of Glycosciences. *Wiley-VCH* **2009,** *28*.
2. Cummings, R. D., The repertoire of glycan determinants in the human glycome. *Mol Biosyst* **2009,** *5* (10), 1087-104.
3. (a) Kiessling, L. L.; Splain, R. A., Chemical approaches to glycobiology. *Annu Rev Biochem* **2010,** *79*, 619-53; (b) Gabius, H. J.; Andre, S.; Jimenez-Barbero, J.; Romero, A.; Solis, D., From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem Sci* **2011,** *36* (6), 298-313.
4. (a) Arnold, J. N.; Saldova, R.; Hamid, U. M.; Rudd, P. M., Evaluation of the serum N-linked glycome for the diagnosis of cancer and chronic inflammation. *Proteomics* **2008,** *8* (16), 3284-93; (b) Drake, P. M.; Cho, W.; Li, B.; Prakobphol, A.; Johansen, E.; Anderson, N. L.; Regnier, F. E.; Gibson, B. W.; Fisher, S. J., Sweetening the pot: adding glycosylation to the biomarker discovery equation. *Clin Chem* **2010,** *56* (2), 223-36; (c) Reis, C. A.; Osorio, H.; Silva, L.; Gomes, C.; David, L., Alterations in glycosylation as biomarkers for cancer detection. *J Clin Pathol* **2010,** *63* (4), 322-9; (d) Gabius, H. J., Glycobiomarkers by glycoproteomics and glycan profiling (glycomics): emergence of functionality. *Biochem Soc Trans* **2011,** *39* (1), 399-405; (e) Amano, M.; Eriksson, H.; Manning, J. C.; Detjen, K. M.; Andre, S.; Nishimura, S.; Lehtio, J.; Gabius, H. J., Tumour suppressor p16(INK4a) - anoikis-favouring decrease in N/O-glycan/cell surface sialylation by down-regulation of enzymes in sialic acid biosynthesis in tandem in a pancreatic carcinoma model. *FEBS J* **2012,** *279* (21), 4062-80.
5. (a) Etxebarria, J.; Calvo, J.; Martin-Lomas, M.; Reichardt, N. C., Lectin-array blotting: profiling protein glycosylation in complex mixtures. *ACS Chem Biol* **2012,** *7* (10), 1729-37; (b) Song, X.; Yu, H.; Chen, X.; Lasanajak, Y.; Tappert, M. M.; Air, G. M.; Tiwari, V. K.; Cao, H.; Chokhawala, H. A.; Zheng, H.; Cummings, R. D.; Smith, D. F., A sialylated glycan microarray reveals novel interactions of modified sialic acids with proteins and viruses. *J Biol Chem* **2011,** *286* (36), 31610-22; (c) Padler-Karavani, V.; Song, X.; Yu, H.; Hurtado-Ziola, N.; Huang, S.; Muthana, S.; Chokhawala, H. A.; Cheng, J.; Verhagen, A.; Langereis, M. A.; Kleene, R.; Schachner, M.; de Groot, R. J.; Lasanajak, Y.; Matsuda, H.; Schwab, R.; Chen, X.; Smith, D. F.; Cummings, R. D.; Varki, A., Cross-comparison of protein recognition of sialic acid diversity on two novel sialoglycan microarrays. *J Biol Chem* **2012,** *287* (27), 22593-608; (d) Blixt, O.; Head, S.; Mondala, T.; Scanlan, C.; Huflejt, M. E.; Alvarez, R.; Bryan, M. C.; Fazio, F.; Calarese, D.; Stevens, J.; Razi, N.; Stevens, D. J.; Skehel, J. J.; van Die, I.; Burton, D. R.; Wilson, I. A.; Cummings, R.; Bovin, N.; Wong, C. H.; Paulson, J. C., Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc Natl Acad Sci U S A* **2004,** *101* (49), 17033-8; (e) Song, X.; Xia, B.; Stowell, S. R.; Lasanajak, Y.; Smith, D. F.; Cummings, R. D., Novel fluorescent glycan microarray strategy reveals ligands for galectins. *Chem Biol* **2009,** *16* (1), 36-47.
6. (a) Andre, S.; Kozar, T.; Kojima, S.; Unverzagt, C.; Gabius, H. J., From structural to functional glycomics: core substitutions as molecular switches for shape and lectin affinity of N-glycans. *Biol Chem* **2009,** *390* (7), 557-65; (b) Andre, S.; Kozar, T.; Schuberth, R.; Unverzagt, C.; Kojima, S.; Gabius, H. J., Substitutions in the N-

glycan core as regulators of biorecognition: the case of core-fucose and bisecting GlcNAc moieties. *Biochemistry* **2007,** *46* (23), 6984-95.
7. Schauer, R., Sialic acids as regulators of molecular and cellular interactions. *Curr Opin Struct Biol* **2009,** *19* (5), 507-14.
8. Goldstein, I. J.; Poretz, R. D., The Lectins: Properties, Functions and Applications in Biology and Medicine. *Academic Press: San Diego, CA* **1986**, 33-247.
9. Jimenez-Barbero, J.; Javier Canada, F.; Asensio, J. L.; Aboitiz, N.; Vidal, P.; Canales, A.; Groves, P.; Gabius, H. J.; Siebert, H. C., Hevein domains: an attractive model to study carbohydrate-protein interactions at atomic resolution. *Adv Carbohydr Chem Biochem* **2006,** *60*, 303-54.
10. (a) Asensio, J. L.; Canada, F. J.; Bruix, M.; Rodriguez-Romero, A.; Jimenez-Barbero, J., The interaction of hevein with N-acetylglucosamine-containing oligosaccharides. Solution structure of hevein complexed to chitobiose. *Eur J Biochem* **1995,** *230* (2), 621-33; (b) Asensio, J. L.; Canada, F. J.; Siebert, H. C.; Laynez, J.; Poveda, A.; Nieto, P. M.; Soedjanaamadja, U. M.; Gabius, H. J.; Jimenez-Barbero, J., Structural basis for chitin recognition by defense proteins: GlcNAc residues are bound in a multivalent fashion by extended binding sites in hevein domains. *Chem Biol* **2000,** *7* (7), 529-43.
11. (a) Andre, S.; Unverzagt, C.; Kojima, S.; Dong, X.; Fink, C.; Kayser, K.; Gabius, H. J., Neoglycoproteins with the synthetic complex biantennary nonasaccharide or its alpha 2,3/alpha 2,6-sialylated derivatives: their preparation, assessment of their ligand properties for purified lectins, for tumor cells in vitro, and in tissue sections, and their biodistribution in tumor-bearing mice. *Bioconjug Chem* **1997,** *8* (6), 845-55; (b) Galanina, O. E.; Kaltner, H.; Khraltsova, L. S.; Bovin, N. V.; Gabius, H. J., Further refinement of the description of the ligand-binding characteristics for the galactoside-binding mistletoe lectin, a plant agglutinin with immunomodulatory potency. *J Mol Recognit* **1997,** *10* (3), 139-47; (c) Unverzagt, C.; Andre, S.; Seifert, J.; Kojima, S.; Fink, C.; Srikrishna, G.; Freeze, H.; Kayser, K.; Gabius, H. J., Structure-activity profiles of complex biantennary glycans with core fucosylation and with/without additional alpha 2,3/alpha 2,6 sialylation: synthesis of neoglycoproteins and their properties in lectin assays, cell binding, and organ uptake. *J Med Chem* **2002,** *45* (2), 478-91.
12. (a) Geisler, C.; Jarvis, D. L., Effective glycoanalysis with Maackia amurensis lectins requires a clear understanding of their binding specificities. *Glycobiology* **2011,** *21* (8), 988-93; (b) Knibbs, R. N.; Goldstein, I. J.; Ratcliffe, R. M.; Shibuya, N., Characterization of the carbohydrate binding specificity of the leukoagglutinating lectin from Maackia amurensis. Comparison with other sialic acid-specific lectins. *J Biol Chem* **1991,** *266* (1), 83-8; (c) Kawaguchi, T.; Matsumoto, I.; Osawa, T., Studies on hemagglutinins from Maackia amurensis seeds. *J Biol Chem* **1974,** *249* (9), 2786-92.
13. Hernandez-Gay, J. J.; Arda, A.; Eller, S.; Mezzato, S.; Leeflang, B. R.; Unverzagt, C.; Canada, F. J.; Jimenez-Barbero, J., Insights into the dynamics and molecular recognition features of glycopeptides by protein receptors: the 3D solution structure of hevein bound to the trisaccharide core of N-glycoproteins. *Chemistry* **2010,** *16* (35), 10715-26.
14. Espinosa, J. F.; Asensio, J. L.; Garcia, J. L.; Laynez, J.; Bruix, M.; Wright, C.; Siebert, H. C.; Gabius, H. J.; Canada, F. J.; Jimenez-Barbero, J., NMR investigations of protein-carbohydrate interactions binding studies and refined three-dimensional solution structure of the complex between the B domain of wheat germ agglutinin and N,N', N''-triacetylchitotriose. *Eur J Biochem* **2000,** *267* (13), 3965-78.
15. (a) Wright, C. S., Structural comparison of the two distinct sugar binding sites in wheat germ agglutinin isolectin II. *J Mol Biol* **1984,** *178* (1), 91-104; (b) Wright, C. S.,

2.2 A resolution structure analysis of two refined N-acetylneuraminyl-lactose--wheat germ agglutinin isolectin complexes. *J Mol Biol* **1990,** *215* (4), 635-51; (c) Wright, C. S., Crystal structure of a wheat germ agglutinin/glycophorin-sialoglycopeptide receptor complex. Structural basis for cooperative lectin-cell binding. *J Biol Chem* **1992,** *267* (20), 14345-52; (d) Wright, C. S.; Jaeger, J., Crystallographic refinement and structure analysis of the complex of wheat germ agglutinin with a bivalent sialoglycopeptide from glycophorin A. *J Mol Biol* **1993,** *232* (2), 620-38.

16. (a) Kronis, K. A.; Carver, J. P., Specificity of isolectins of wheat germ agglutinin for sialyloligosaccharides: a 360-MHz proton nuclear magnetic resonance binding study. *Biochemistry* **1982,** *21* (13), 3050-7; (b) Kronis, K. A.; Carver, J. P., Wheat germ agglutinin dimers bind sialyloligosaccharides at four sites in solution: proton nuclear magnetic resonance temperature studies at 360 MHz. *Biochemistry* **1985,** *24* (4), 826-33; (c) Wang, W. C.; Cummings, R. D., The immobilized leukoagglutinin from the seeds of Maackia amurensis binds with high affinity to complex-type Asn-linked oligosaccharides containing terminal sialic acid-linked alpha-2,3 to penultimate galactose residues. *J Biol Chem* **1988,** *263* (10), 4576-85.

17. Jimenez, M.; Andre, S.; Siebert, H. C.; Gabius, H. J.; Solis, D., AB-type lectin (toxin/agglutinin) from mistletoe: differences in affinity of the two galactoside-binding Trp/Tyr-sites and regulation of their functionality by monomer/dimer equilibrium. *Glycobiology* **2006,** *16* (10), 926-37.

18. (a) Jimenez, M.; Andre, S.; Barillari, C.; Romero, A.; Rognan, D.; Gabius, H. J.; Solis, D., Domain versatility in plant AB-toxins: evidence for a local, pH-dependent rearrangement in the 2gamma lectin site of the mistletoe lectin by applying ligand derivatives and modelling. *FEBS Lett* **2008,** *582* (15), 2309-12; (b) Mikeska, R.; Wacker, R.; Arni, R.; Singh, T. P.; Mikhailov, A.; Gabdoulkhakov, A.; Voelter, W.; Betzel, C., Mistletoe lectin I in complex with galactose and lactose reveals distinct sugar-binding properties. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **2005,** *61* (Pt 1), 17-25.

19. Muthing, J.; Meisen, I.; Bulau, P.; Langer, M.; Witthohn, K.; Lentzen, H.; Neumann, U.; Peter-Katalinic, J., Mistletoe lectin I is a sialic acid-specific lectin with strict preference to gangliosides and glycoproteins with terminal Neu5Ac alpha 2-6Gal beta 1-4GlcNAc residues. *Biochemistry* **2004,** *43* (11), 2996-3007.

20. Konami, Y.; Yamamoto, K.; Osawa, T.; Irimura, T., Strong affinity of Maackia amurensis hemagglutinin (MAH) for sialic acid-containing Ser/Thr-linked carbohydrate chains of N-terminal octapeptides from human glycophorin A. *FEBS Lett* **1994,** *342* (3), 334-8.

21. Imberty, A.; Gautier, C.; Lescar, J.; Perez, S.; Wyns, L.; Loris, R., An unusual carbohydrate binding site revealed by the structures of two Maackia amurensis lectins complexed with sialic acid-containing oligosaccharides. *J Biol Chem* **2000,** *275* (23), 17541-8.

22. (a) Andre, S.; Sanchez-Ruderisch, H.; Nakagawa, H.; Buchholz, M.; Kopitz, J.; Forberich, P.; Kemmner, W.; Bock, C.; Deguchi, K.; Detjen, K. M.; Wiedenmann, B.; von Knebel Doeberitz, M.; Gress, T. M.; Nishimura, S.; Rosewicz, S.; Gabius, H. J., Tumor suppressor p16INK4a--modulator of glycomic profile and galectin-1 expression to increase susceptibility to carbohydrate-dependent induction of anoikis in pancreatic carcinoma cells. *FEBS J* **2007,** *274* (13), 3233-56; (b) Patsos, G.; Andre, S.; Roeckel, N.; Gromes, R.; Gebert, J.; Kopitz, J.; Gabius, H. J., Compensation of loss of protein function in microsatellite-unstable colon cancer cells (HCT116): a gene-dependent effect on the cell surface glycan profile. *Glycobiology* **2009,** *19* (7), 726-34; (c) Van de Wouwer, M.; Andre, S.; Gabius, H. J.; Villalobo, A., Nitric oxide changes distinct

aspects of the glycophenotype of human neuroblastoma NB69 cells. *Nitric Oxide* **2011,** *24* (2), 91-101; (d) Varki, A.; Gagneux, P., Multifarious roles of sialic acids in immunity. *Ann N Y Acad Sci* **2012,** *1253*, 16-36.
23. (a) Martin-Santamaria, S.; Andre, S.; Buzamet, E.; Caraballo, R.; Fernandez-Cureses, G.; Morando, M.; Ribeiro, J. P.; Ramirez-Gualito, K.; de Pascual-Teresa, B.; Canada, F. J.; Menendez, M.; Ramstrom, O.; Jimenez-Barbero, J.; Solis, D.; Gabius, H. J., Symmetric dithiodigalactoside: strategic combination of binding studies and detection of selectivity between a plant toxin and human lectins. *Org Biomol Chem* **2011,** *9* (15), 5445-55; (b) Mayer, M.; Meyer, B., Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew Chem Int Ed Engl* **1999,** *38* (12), 1784-1788; (c) Jimenez-Barbero, J.; Dragoni, E.; Venturi, C.; Nannucci, F.; Arda, A.; Fontanella, M.; Andre, S.; Canada, F. J.; Gabius, H. J.; Nativi, C., Alpha-O-linked glycopeptide mimetics: synthesis, conformation analysis, and interactions with viscumin, a galactoside-binding model lectin. *Chemistry* **2009,** *15* (40), 10423-31.
24. Kaminski, G.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L., Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001,** *105*, 6474-6487.
25. Maestro, A Powerful, All-Purpose Molecular Modeling Environment, Version 8.5. *Schroedinger, LLC* **2008,** *New York*.
26. Gabius, H. J.; Darro, F.; Remmelink, M.; Andre, S.; Kopitz, J.; Danguy, A.; Gabius, S.; Salmon, I.; Kiss, R., Evidence for stimulation of tumor proliferation in cell lines and histotypic cultures by clinically relevant low doses of the galactoside-binding mistletoe lectin, a component of proprietary extracts. *Cancer Invest* **2001,** *19* (2), 114-6.
27. Kajihara, Y.; Suzuki, Y.; Yamamoto, N.; Sasaki, K.; Sakakibara, T.; Juneja, L. R., Prompt chemoenzymatic synthesis of diverse complex-type oligosaccharides and its application to the solid-phase synthesis of a glycopeptide with Asn-linked sialyl-undeca- and asialo-nonasaccharides. *Chemistry* **2004,** *10* (4), 971-85.
28. Seko, A.; Koketsu, M.; Nishizono, M.; Enoki, Y.; Ibrahim, H. R.; Juneja, L. R.; Kim, M.; Yamamoto, T., Occurence of a sialylglycopeptide and free sialylglycans in hen's egg yolk. *Biochim Biophys Acta* **1997,** *1335* (1-2), 23-32.

Chapter 4. The recognition of galacto-oligosaccharides by specific scFv antibody fragments by using STD-NMR spectroscopy

4.1 Introduction

The immunological importance of the non-human galactosyl-α-(1→3)-galactose (Gal-α-(1→3)-Gal) disaccharide epitope was initially highlighted by its direct involvement in the hyperacute rejection (HAR) of porcine organ xenografts[1]. Xenotransplantation has been widely considered as a direct means of overcoming the critical shortage of human donor organs. Porcine organs, in particular, were first considered as the most suitable replacements[2]. However, porcine tissue expresses a high proportion of the Gal-α-(1→3)-Gal epitope and elicits a vigorous anti-Gal-α-(1→3)-Gal antibody response, following transplantation of porcine xenografts into humans[3].

The Galα1→3Gal disaccharide is found at the non reducing terminus of protein- and lipid-bound glycans, and the α-linked galactose is a non-charged alternative to sialic acid as a chain terminator (Fig. 4.1). Galα(1→3)Gal structures have been detected in a varietyof mammalian species. Only cells and tissues derived from humans, apes and Old World monkeys do not express the α-gal glycotope[4], due to the inactivation of the α1→3-Galactosyltransferase (α1→3-GT) biosynthesis route[5].Therefore, in these organisms the sialylated version is the only one that exists, whereas in the rest of species, the galactosylated form still takes place (See Fig 4.1a and 4.1b).

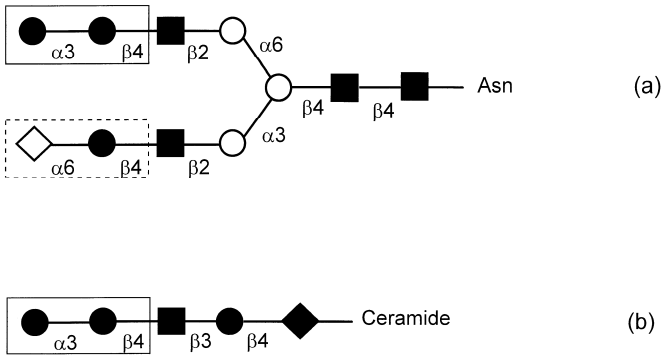


Figure 4.1 View of the terminal structures of glycoprotein and glycolipidglycans[1]. (a) A diantennary *N*-glycan terminated by sialic acid and α1,3-linked galactose. (b) The ceramide-pentahexoside.

However, the α-gal glycotope is found in tissue cultures both on mammalian tissues and cells. In fact, in the mouse, *G. simplicifolia* I-B4 (GSI-B4) lectin staining has been employed to detect the Galα(1→3)Gal structure in a variety of organs and cell types, with the exception of liver[6]. Specially, high-level expression is found on the endothelial cells of all capillaries, arterioles and arteries, the first point of contact between antibodies and donor organ.

Cairns et al.[7]and Vaughan et al.[8] have detected both Galα(1→3)Gal containing glycoproteins and glycosphingolipids on pig vascular endothelium. The detailed structural analyses of the xeno-epitope was carried out by Samuelsson etal[9], who showed that the major α(1→3)-galactosylated glycosphingolipid in the porcine endothelium is the ceramide pentasaccharide [Galα(1→3)Galβ(1→4)GlcNAcβ(1→3)-Galβ(1→4)Glc-ceramide, Fig. 4.1b). The protein-linked glycans contains the Galα(1→3)Galβ1→4GlcNAc trisaccharide at their non-reducing terminus (Fig. 4.1).

Other relevant studies, on the characterization of anti-Gal antibodies using thin layer chromatography (TLC) demonstrated that anti-Gal recognizes glycosphingolipids with non-reducing terminal Galα**1→3**Galβ1→4GlcNAc-R structures isolated from rabbit red blood cells, whereas it does not bind to a range of closely related glycosphingolipids (as glycosphingolipids with the Galα**1→4**Galβ1→4GlcNAc-R structures). This fact establishes that anti-Gal is able to distinguish between structures that have the same carbohydrate sequence and anomeric configuration, but which differ in the linkage positions of the terminal Gal residues[10].

Undesirable immune responses to this Galα1→3Galβ1→4GlcNAc-R epitope are also thought to be involved in the suboptimal clinical outcomes after implantation of acellular tissue matrices[11]. Naturally occurring human antibodies directed against the Gal-α-(1→3)-Gal epitope are widespread in humans[10], comprising up to 3% of immunoglobulin (Ig) in human sera (mainly the IgG2 subclass)[3b, c]. Furthermore, the possibility of exploiting the Gal-α-(1,3)-Gal epitope/natural human antibody recognition system to improve the efficacy of autologous vaccines is recently gaining increased attention [12].

4.2 Objectives

Herein we describe the analysis of the recognition features of selected ligands by specific scFv G12, A4 and A11 antibody fragments. STD NMR methods have been employed to access residue-specific binding information. The STD analysis were applied to monitor the interactions ofα-3-O-*Galactobiose* (or Glycotope), the linear-B-trisaccharide, 3α-4β-*D-Galactotriose* and 3α-4β-3α-*D-Galactotetraose* with the above

mentioned specific antibodies to study the different involvement of each residue in the molecular recognition process.

4.3 The bound state; STD experiments

4.3.1 The recognition of α-3-O-Galactobiose (Glycotope)

STD-NMR experiments were performed to obtain structural information about the epitope of the α-3-O-Galactobiose recognized by the scFv G12, A4 and A11 specific antibody fragments. As shown in Figure 4.2, the STD experiments demonstrated the interaction of scFv G12 with the ligand. Interestingly, STD signals for Gal A' protons (See 1A', 2A', 3A' and 4A' proton signals in Fig. 4.2a) were clearly more intense than those for residues A or B in the presence of scFv G12 (See 4A and 5A proton signals in Figure 4.2a). The deduced average STD values per residue indicated that Gal A' was closer to the binding site.

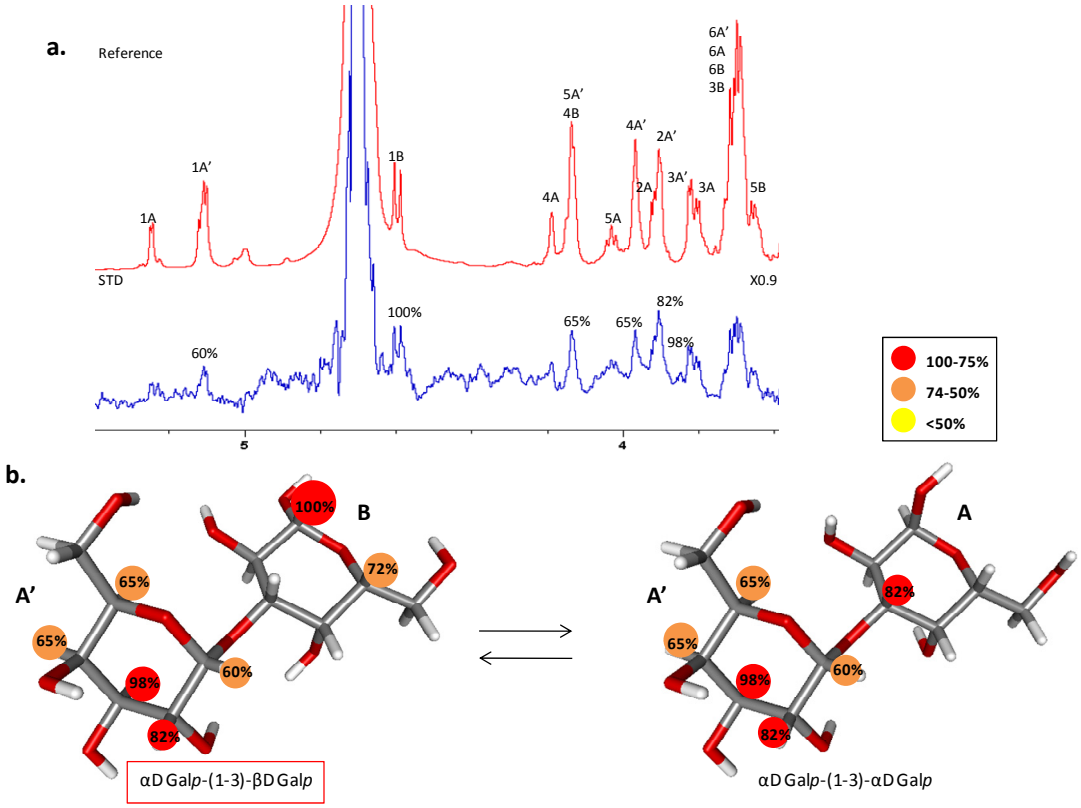


Figure 4.2 The interaction of α-3-O-Galactobiose with the scFv G12 antibody as deduced from the STD experiments for a 90:1 molar ratio. (a) Reference (top) and STD spectra (bottom) with the relevant proton signals highlighted. (b) Schematic view of the 3D structure of α-3-O-Galactobiose with the measured values of the relative STD intensities detected for each proton.

In the presence of scFv A4, again the STD signals of H1-H4 protons of GalA' (Fig. 4.3a) were significantly stronger than those of the other residues. Thus, the analysis of the STD intensities permitted to suggest that the A' residue is the major epitope for scFv A4.

In the presence of scFv A11, the measured STDs were significantly weaker, strongly suggesting the presence of a weaker interaction. Nevertheless, the inspection of the STD values demonstrated again the preferred involvement of GalA' residue in the binding event (See Figure 4.4b).

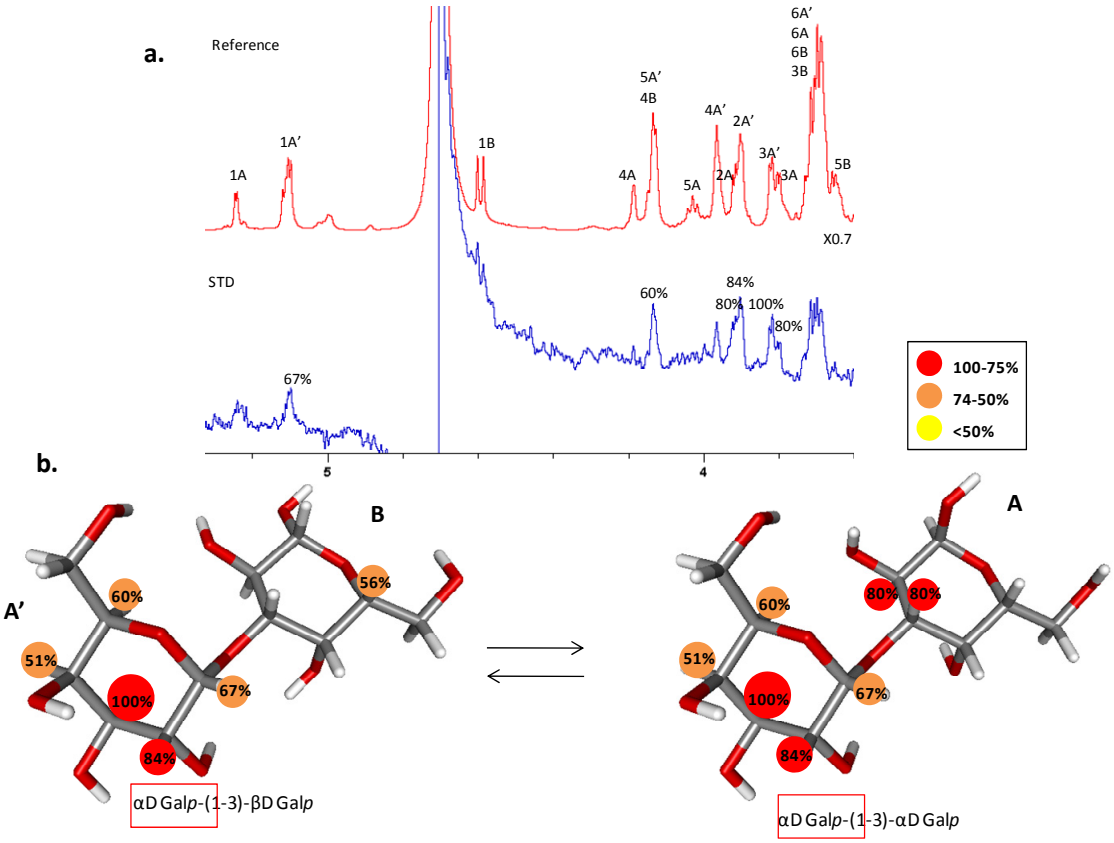


Figure 4.3 The interaction of α-3-O-Galactobiose with the scFv A4 antibody, as deduced from the STD experiments for a 90:1molar ratio. (a) Reference (top) and STD spectra (bottom), with the relevant proton signals highlighted. (b) Schematic view of the 3D structure of α-3-O-Galactobiose with the values of the relative STD intensities detected for each proton.

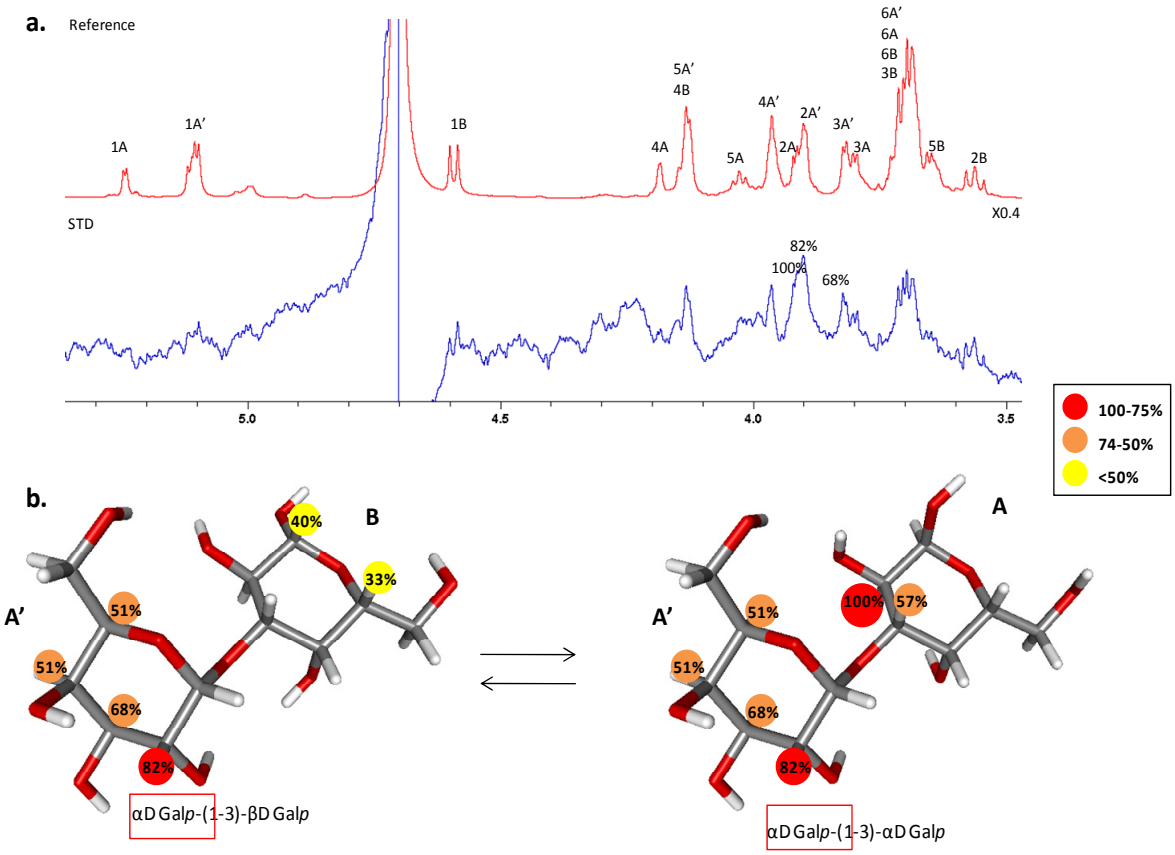


Figure 4.4 The interaction of α-3-O-Galactobiose with scFv A11 antibody, as deduced from the STD experiments for a 90:1molar ratio. (a) Reference (top) and STD spectra (bottom), with the relevant proton signals highlighted. (b) Schematic view of the 3D structure of α-3-O-Galactobiose with the values of the relative STD intensities detected for each proton.

Therefore, the superimposition of the STD spectra of the α-3-O-Galactobiose ligand in the presence of either scFv A11, A4 and G12 antibody fragments showed the existence of a similar STD effect pattern, thus indicating that the ligand epitope recognized by the three specific antibodies is fairly analogous (See protons in red, Figure 4.5).

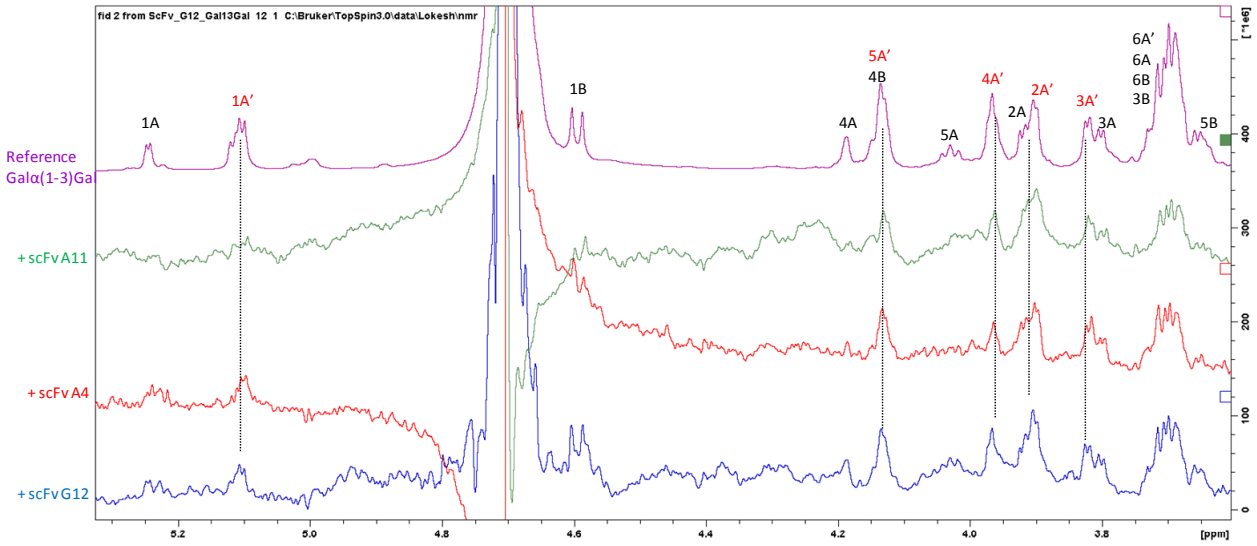


Figure 4.5 Superimposition of the STD spectra obtained for α-3-O-Galactobiose in the presence of both scFv A11 (green), scFv A4 (red) and scFv G12 (blue). The proton signals with the strongest STD effects are shown in red. The reference spectrum is shown in violet.

4.3.2 The relevance of different substitutions: GlcNAc-, Gal- and α-3-O-Galactobiose-containing decorations at the galactobiose scaffold

As a second step, we paid attention to the relevance and the implication of different substitutions at the Galα(1-3)Gal disaccharide scaffold (or glycotope) for the recognition by specific scFv antibody fragments. Thus, STD experiments were then performed with elongated trisaccharides (with a β(1-4)GlcNAc moiety, a B-trisaccharide moiety, and a β(1-4)Gal substitution at the reducing-end, respectively). The corresponding tetrasaccharide (3α-4β-3α-*D-Galactoteraose*), containing twice the glycotope motif, was also studied.

In fact, the STD analysis of the B-trisaccharide in the presence of G12, A4 and A11 clearly revealed the existence of binding. Inspection of Figures 4.6, 4.7 and 4.8 permitted to assess that the strongest STD effect arose from the Gal A' protons (see H2A', H3A' and H4A' protons in Fig. 4.6-4.8) in the presence of the three scFv antibody fragments. The STD effect decreased for the Gal B' and especially for the GlcNAcA residues (see the STD relative intensities for 4B', 4B and NHAcA/B proton signals in Figs 4.6-4.8). Consequently, these residues are less implicated in the binding event, with the GlcNAc moiety point probably towards the solvent.

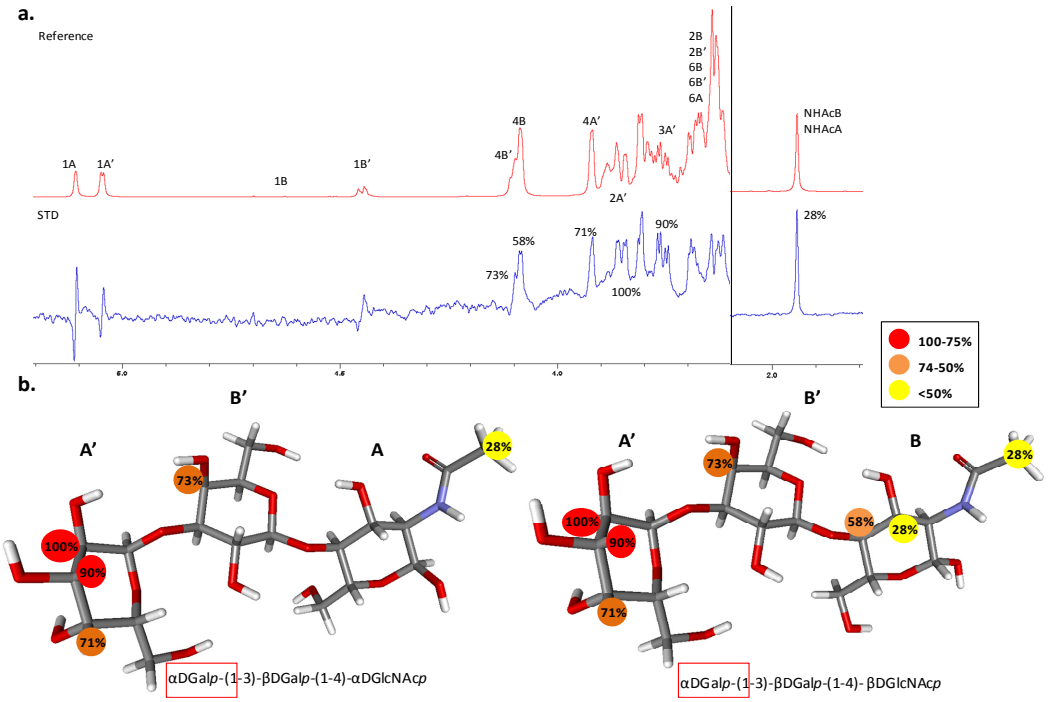


Figure 4.6 The interaction of the B-trisaccharide with the scFv G12 antibody, as deduced from STD experiments for a 90:1 molar ratio. (a) Reference (top) and STD spectra (bottom) with relevant proton signals highlighted. (b) Schematic view of the 3D structure of the B-trisaccharide with the values of the relative STD intensities detected for each proton.

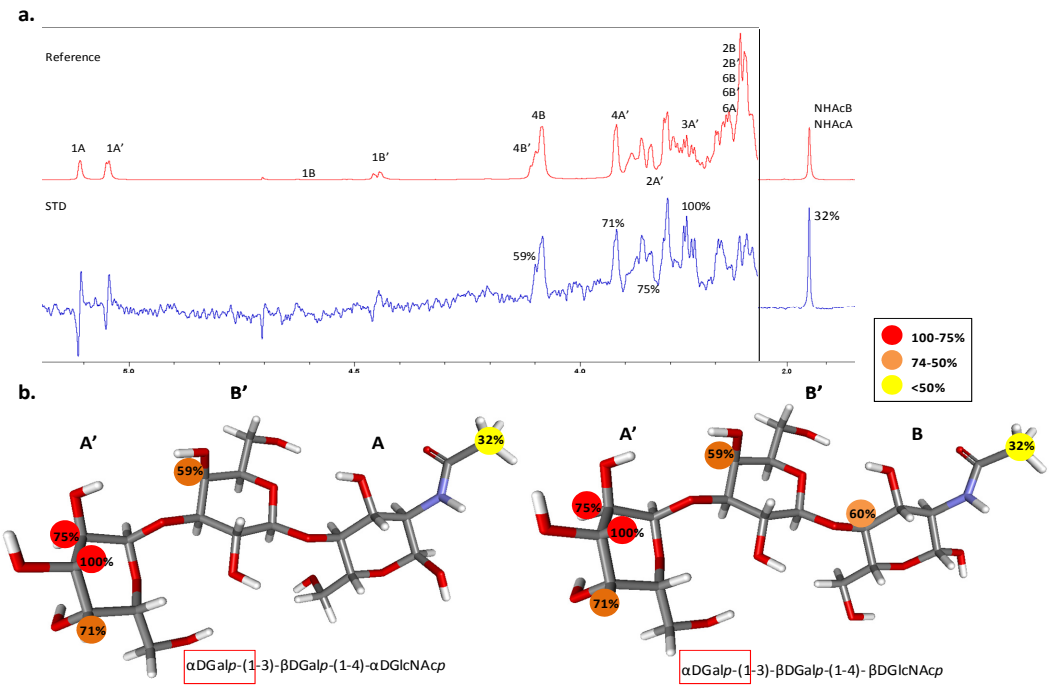


Figure 4.7 The interaction of the B-trisaccharide with scFv A4 antibody, as deduced from STD experiments for a 90:1 molar ratio. (a) Reference (top) and STD spectra (bottom) with relevant proton signals highlighted. (b) Schematic view of the 3D structure of the B-trisaccharide with the values of the relative STD intensities detected for each proton.

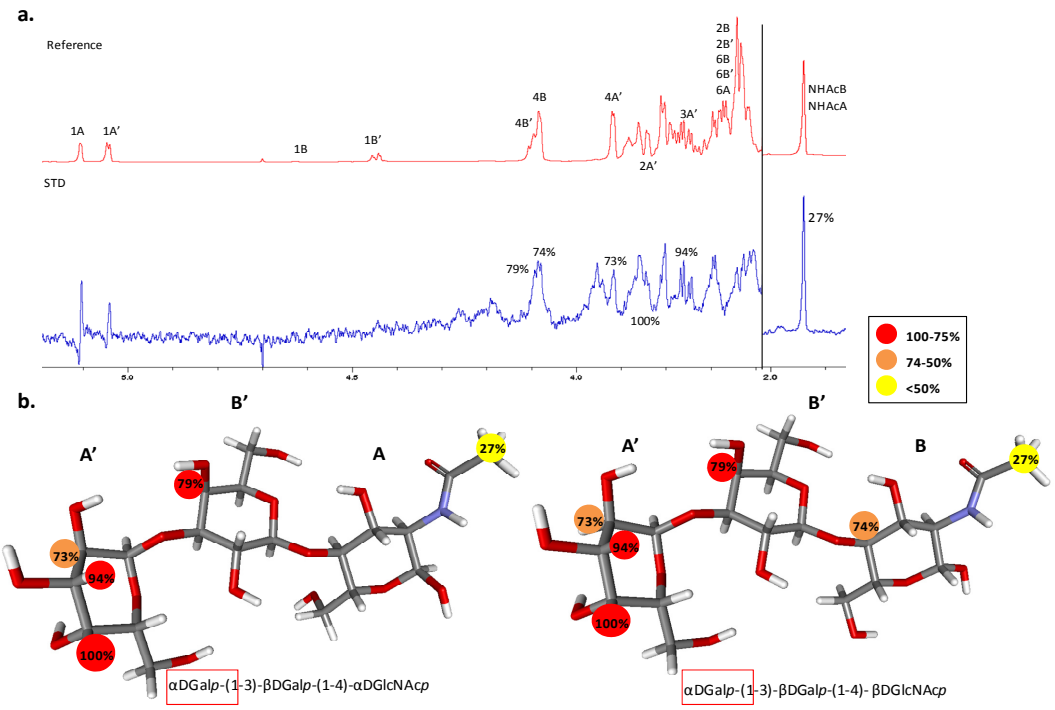


Figure 4.8 The interaction of B-trisaccharide with scFv A11 antibody, as deduced from STD experiments for a 90:1 molar ratio.(a) Reference (top) and STD spectra (bottom) with relevant proton signals highlighted. (b) Schematic view of the 3D structure of the B-trisaccharide with the values of the relative STD intensities detected for each proton.

As described above for the glycotope, the superimposition of the STD spectra of the B-trisaccharide ligand in the presence of either scFv A11, A4 and G12 antibody fragments showed the existence of a similar STD effect pattern, thus indicating that the ligand epitope recognized by the three specific antibodies is fairly similar (See protons in red, Figure 4.9).

STD-NMR experiments were also performed to obtain further information about the presence of a second βGal residue at the reducing end (3α-4β-*D-Galactotriose*). The STD –based analysis definitely demonstrated its recognition by the three scFv G12, A4 and A11 specific antibody fragments. The detailed STD intensity analysis of the signals showed once again the importance of residue Gal A' for binding (see the A' proton signals in Figs 4.10-4.12 and 4.14).

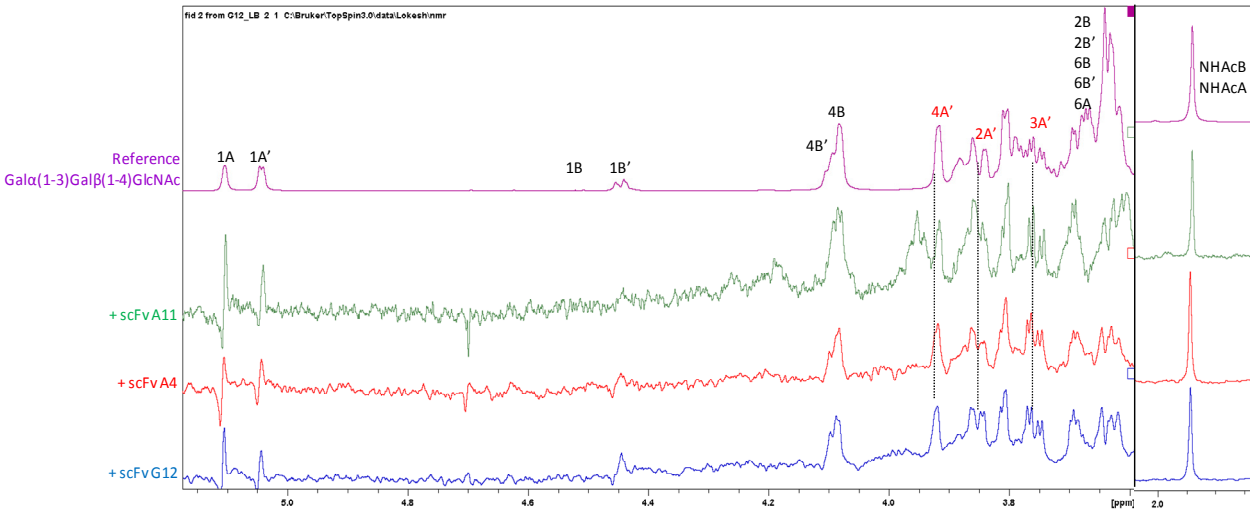


Figure 4.9 Superimposition of the STD spectra obtained for the B-trisaccharide in the presence of either scFv A11 (green), scFv A4 (red) and scFv G12 (blue).The proton signals with strongest STD effects are shown in red. The reference spectrum is shown in violet.

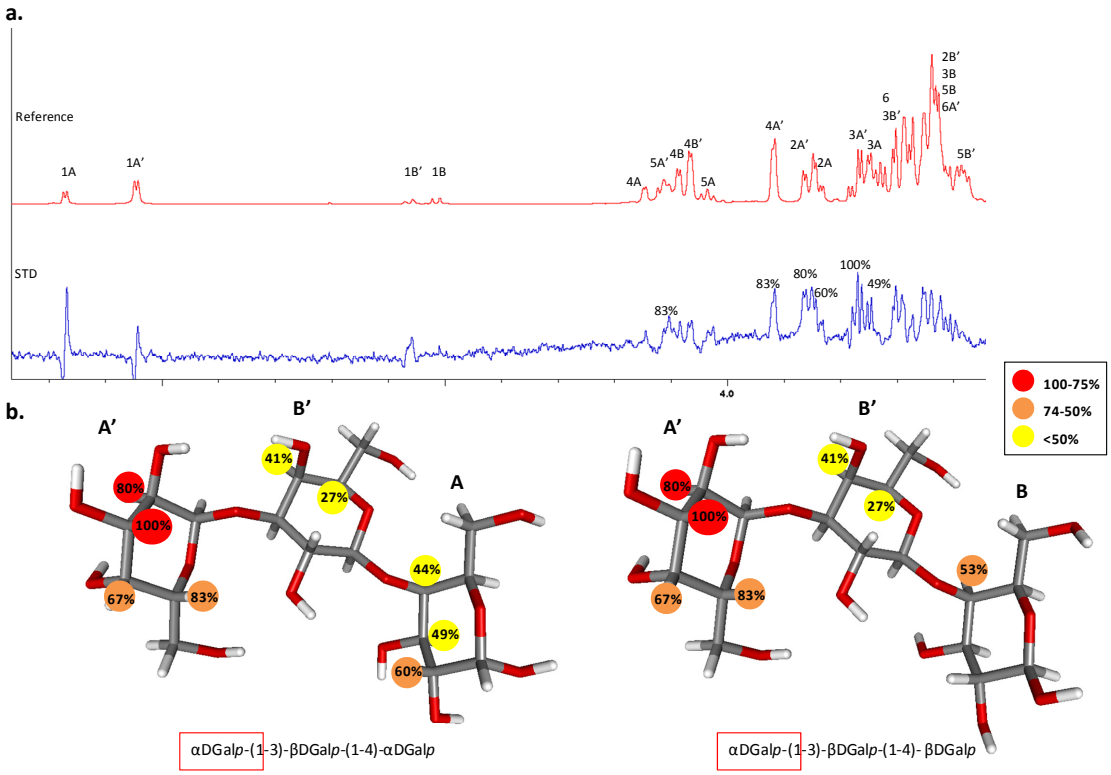


Figure 4.10 The interaction of 3α-4β-*D-Galactotriose* with scFv G12 antibody, as deduced from STD experiments for a 90:1 molar ratio. (a) Reference (top) and STD spectra (bottom) with relevant proton signals highlighted. (b) Schematic view of the 3D structure of 3α-4β-*D-Galactotriose* with the values of the relative STD intensities detected for each proton.

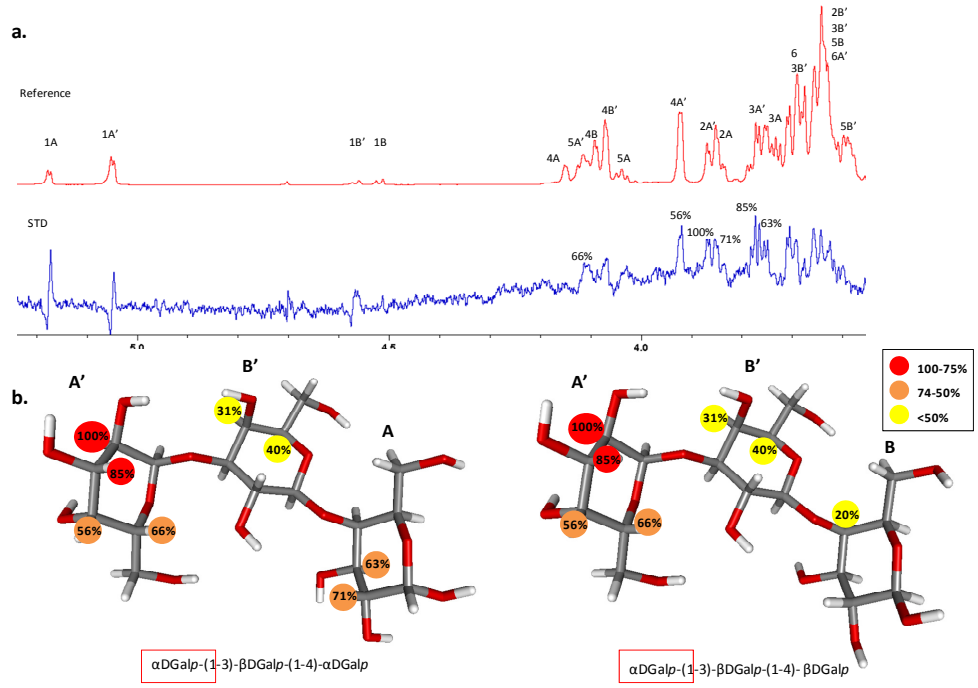


Figure 4.11 The interaction of 3α-4β-*D-Galactotriose* with scFv A4 antibody, as deduced from STD experiments for a 90:1 molar ratio. (a) Reference (top) and STD spectra (bottom) with relevant proton signals highlighted. (b) Schematic view of the 3D structure of 3α-4β-*D-Galactotriose* with the values of the relative STD intensities detected for each proton.

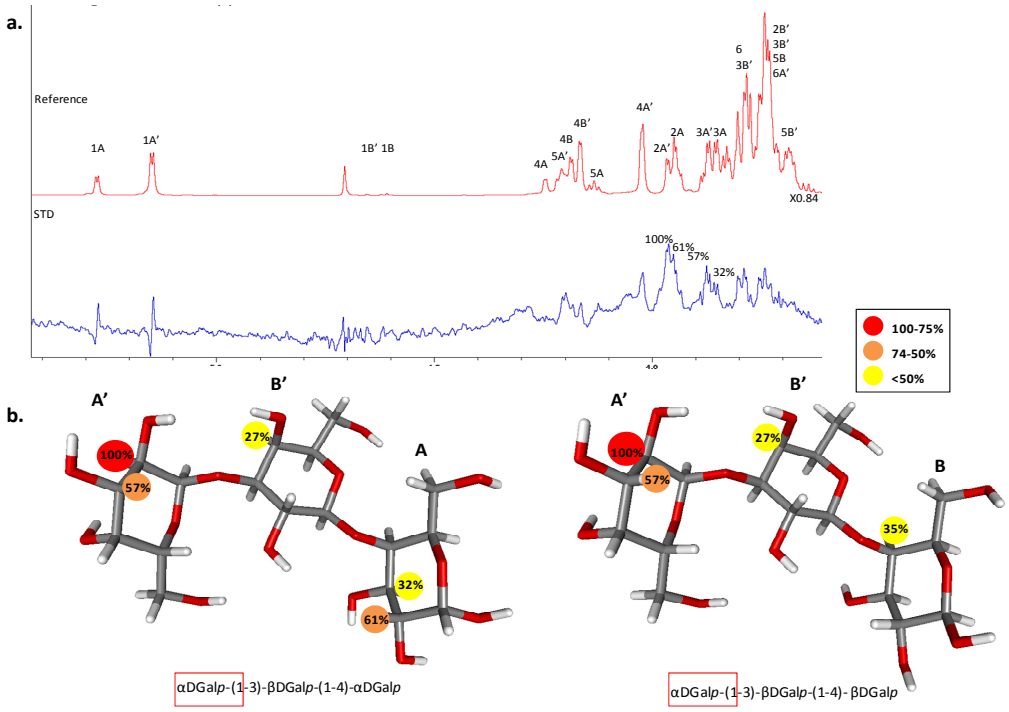


Figure 4.12 The interaction of 3α-4β-*D-Galactotriose* with scFv A11 antibody, as deduced from STD experiments for a 90:1 molar ratio. (a) Reference (top) and STD spectra (bottom) with relevant proton signals highlighted. (b) Schematic view of the 3D structure of 3α-4β-*D-Galactotriose* with the values of the relative STD intensities detected for each proton.

To verify the key involvement of this residue for the recognition process, we also carried out a similar STD study, but now employing the reducing end disaccharide moiety (4β-*D*-*Galactobiose*, that is, devoid of the α-*D*-Gal residue at the non-reducing end) in the presence of scFv A11 antibody fragment using the same conditions for the experiment. Fittingly (Figure 4.13), no STD effect was observed at all, thus demonstrating the essential involvement of the Gal A' residue in the molecular recognition process.

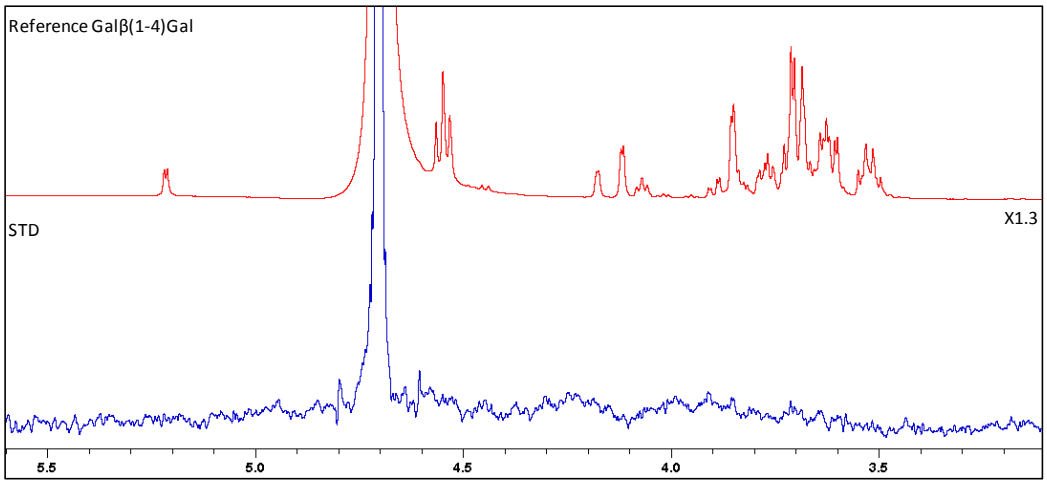


Figure 4.13 STD experiments of 4β-*D*-*Galactobiose*in the presence of scFv A11 antibody with a molar ratio 90:1. Reference (top) and STD spectra (bottom). No STD effect was observed.

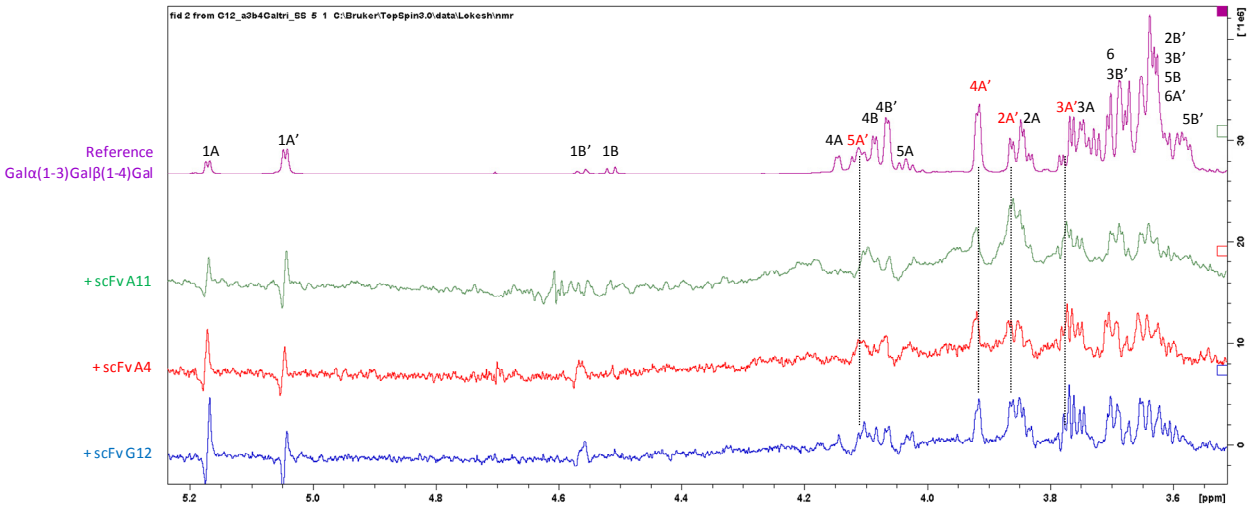


Figure 4.14 Superimposition of the STD spectra obtained for the 3α-4β-*D*-*Galactotriose* in the presence of either scFv A11 (green), scFv A4 (red) and scFv G12 (blue). The proton signals with strongest STD effects are shown in red. The reference spectrum is shown in violet.

Finally, a model tetrasaccharide was also employed with two α-3-O-Galactobiose units joined by a β(1-4) linkage (3α-4β-3α-*D-Galactotetraose)*.Now, it was clearly demonstrated that the non-terminal galactose A'' moiety is that displaying the strongest STD effects in the presence of all antibody fragments (Figures 4.15-4.18). Interestingly, the inner Gal A' residue protons also showed relatively high STD intensities, suggesting the existence of an extended binding site for this type of linkage or the presence of two distinct binding modes in which either Gal A" or Gal A' are interacting with the antibodies.

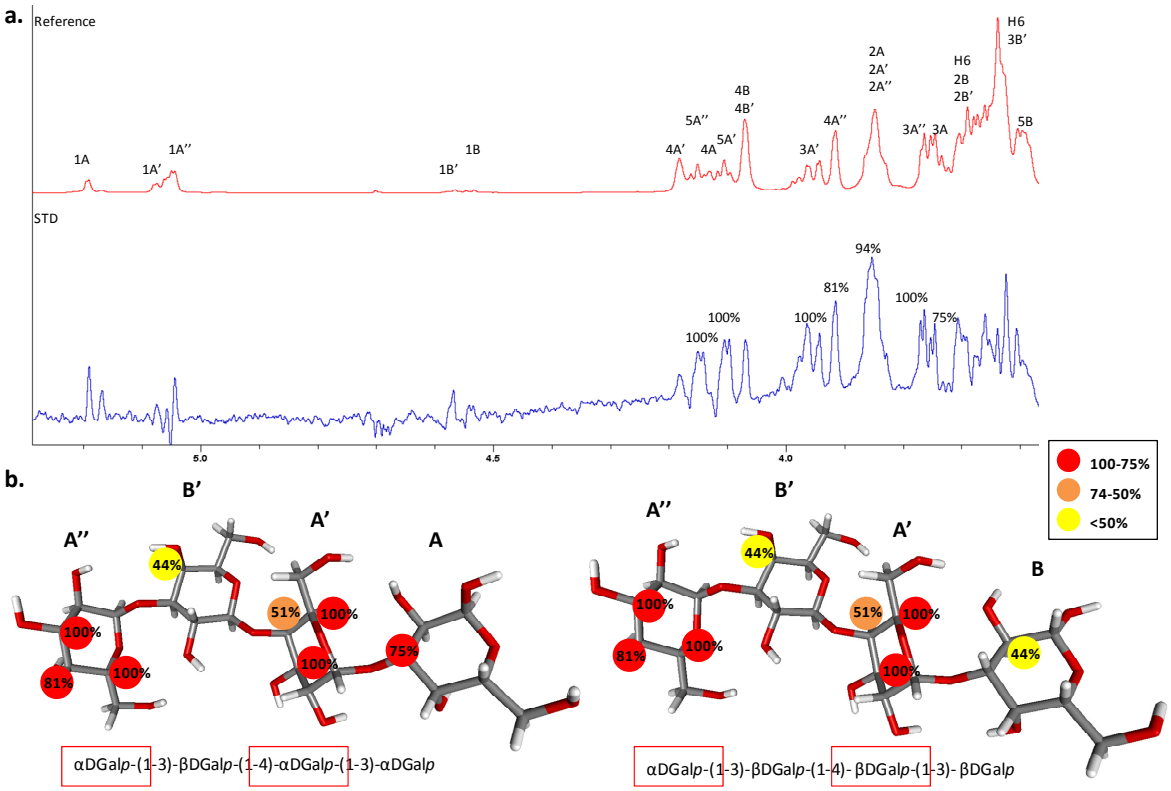


Figure 4.15 The interaction of 3α-4β-3α-*D-Galactotetraose* with the scFv G12 antibody, as deduced from STD experiments for a 90:1 molar ratio. (a) Reference (top) and STD spectra (bottom) with relevant proton signals highlighted. (b) Schematic view of the 3D structure of 3α-4β-3α-*D-Galactotetraose* with the values of the relative STD intensities detected for each proton.

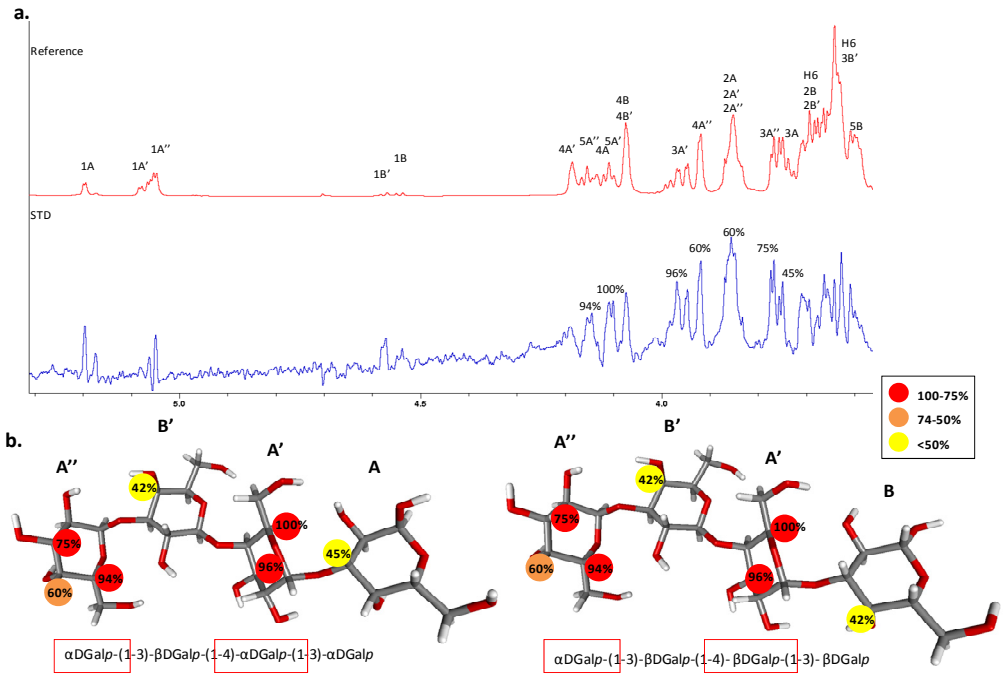


Figure 4.16 The interaction of 3α-4β-3α-*D-Galactotetraose* with the scFv A4 antibody, as deduced from STD experiments for a 90:1 molar ratio. (a) Reference (top) and STD spectra (bottom) with relevant proton signals highlighted. (b) Schematic view of the 3D structure of 3α-4β-3α-*D-Galactotetraose* with the values of the relative STD intensities detected for each proton.

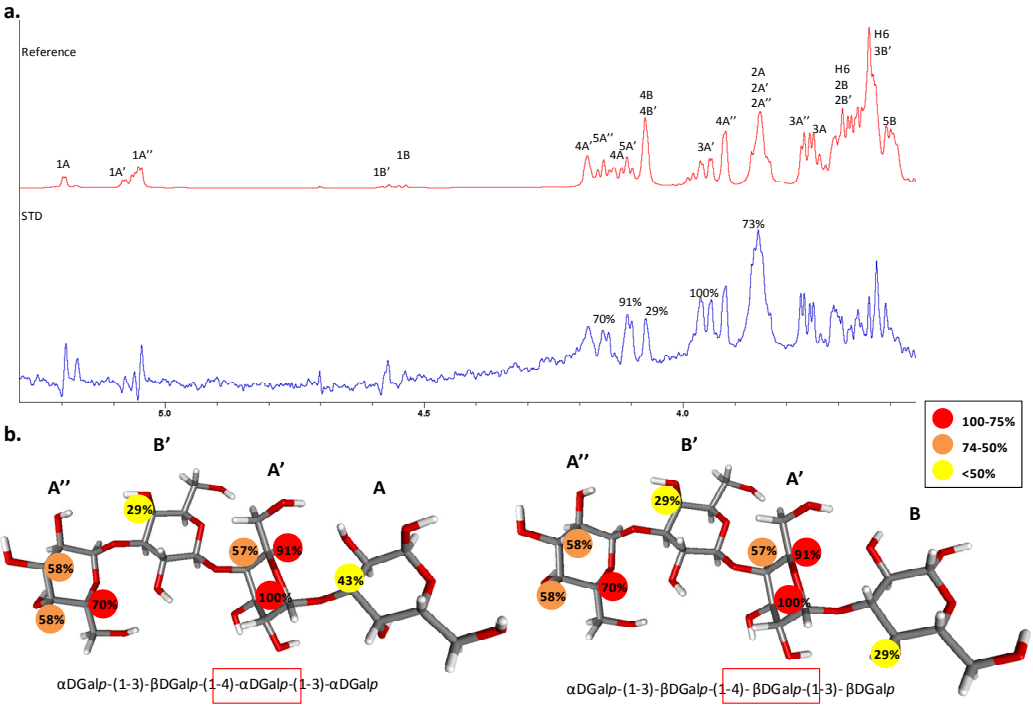


Figure 4.17 The interaction of 3α-4β-3α-*D-Galactotetraose* with the scFv A11 antibody, as deduced from STD experiments for a 90:1 molar ratio. (a) Reference (top) and STD spectra (bottom) with relevant proton signals highlighted. (b) Schematic view of the 3D structure of 3α-4β-3α-*D-Galactotetraose* with the values of the relative STD intensities detected for each proton.

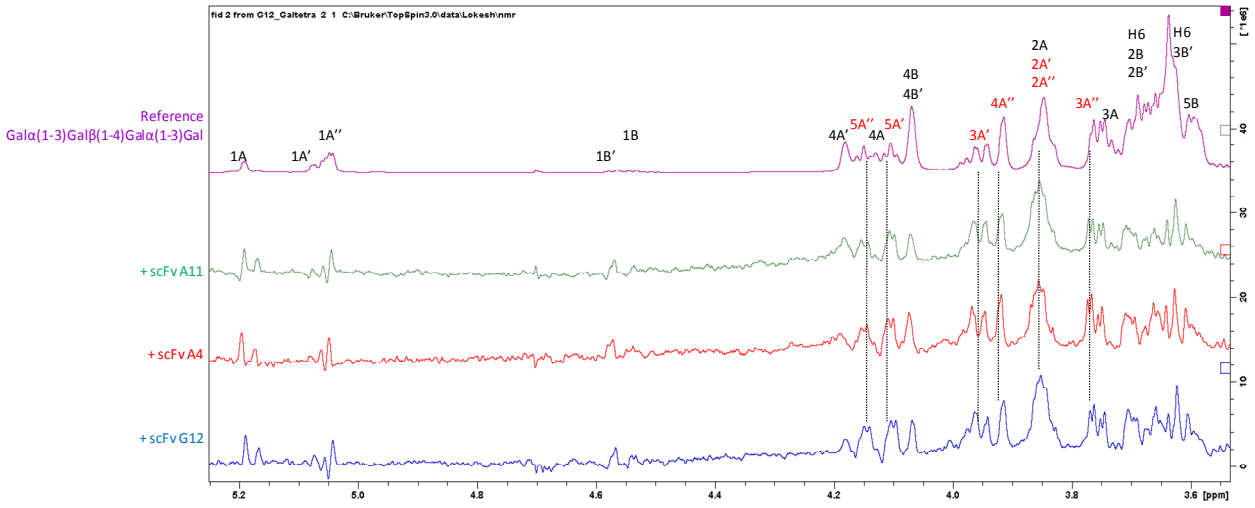


Figure 4.18 Superimposition of the STD spectra obtained for the 3α-4β-3α-*D-Galactotetraose* in the presence of eithers cFv A11 (green), scFv A4 (red) and scFv G12 (blue). The proton signals with strongest STD effects are shown in red. The reference spectrum is shown in violet.

4.4 Conclusions

Four oligosaccharides related to the glycotope antigen have been shown to interact with different scFv glycotope-binding antibody fragments. Using STD experiments, we have shown that the presence of 3-O-α-*DGalactose* A' moiety is a decisive feature for the binding process to take place. In contrast, the other residues at the "reducing end" merely display transient interactions with the carbohydrate binding site.

4.5 Materials and Methods

Methods

Sample preparation and NMR studies

The NMR study of 3α-*D-Galactobiose* (Carbosynth Product Code: OG06491), 4β-*D-Galactobiose* (Carbosynth Product Code: OG15071), 3α-4β-*D-Galactotriose* (Carbosynth Product Code: OG00715), Linear-B-trisaccharide (Carbosynth Product Code: OL06495), and 3α-4β-3α-*D-Galactotetraose* (Carbosynth Product Code: OG05062) oligosaccharides in the free

state was performed using 1mM samples in D_2O. Standard 2D-methods, such as TOCSY (with mixing times of 30 and 80ms), ROESY (300 ms), NOESY (600 ms), and HSQC experiments were employed, using Bruker Avance 500 and 600 MHz (equipped with a triple channel cryoprobe) NMR spectrometers. The target scFv antibody fragments G12, A4 and A11 were prepared at The National University of Ireland (NUI Galway). Samples were dialyzed with phosphate-buffered saline in deuterated water (PBS, pH= 7.3, uncorrected) for the ligand binding experiments.

Saturation transfer difference (STD)[13] experiments were performed on both 500 MHz and 600 MHz NMR spectrometers, at 298K or 310K. The experiments were performed using samples with a scFv antibody fragment/ligand molar ratio of 1/90. The employed concentrations were 25uM/2.25mM for experiments with scFv G12, A4 and A11 antibody fragments. Different experiments (8000 scans each) were performed in which the saturation time of the protein was 2s. Selective "on-resonance" irradiation of antibody resonances was performed at δ0.8ppm, while the "off-resonance" frequency was always set at 100ppm. STD experiments on free ligands at 1mM in D_2O were also acquired in order to obtain "blank" experiments.

4.6 References

1. Joziasse, D. H.; Oriol, R., Xenotransplantation: the importance of the Galalpha1,3Gal epitope in hyperacute vascular rejection. *Biochim Biophys Acta* **1999,** *1455* (2-3), 403-18.
2. (a) Halperin, E. C., Non-human to human organ transplantation: its biologic basis and a potential role for radiation therapy. *Int J Cancer* **2001,** *96* (2), 76-89; (b) Thompson, P.; Badell, I. R.; Lowe, M.; Turner, A.; Cano, J.; Avila, J.; Azimzadeh, A.; Cheng, X.; Pierson, R. N., 3rd; Johnson, B.; Robertson, J.; Song, M.; Leopardi, F.; Strobert, E.; Korbutt, G.; Rayat, G.; Rajotte, R.; Larsen, C. P.; Kirk, A. D., Alternative immunomodulatory strategies for xenotransplantation: CD40/154 pathway-sparing regimens promote xenograft survival. *Am J Transplant* **2012,** *12* (7), 1765-75.
3. (a) Chen, G.; Qian, H.; Starzl, T.; Sun, H.; Garcia, B.; Wang, X.; Wise, Y.; Liu, Y.; Xiang, Y.; Copeman, L.; Liu, W.; Jevnikar, A.; Wall, W.; Cooper, D. K.; Murase, N.; Dai, Y.; Wang, W.; Xiong, Y.; White, D. J.; Zhong, R., Acute rejection is associated with antibodies to non-Gal antigens in baboons using Gal-knockout pig kidneys. *Nat Med* **2005,** *11* (12), 1295-8; (b) Yu, P. B.; Holzknecht, Z. E.; Bruno, D.; Parker, W.; Platt, J. L., Modulation of natural IgM binding and complement activation by natural IgG antibodies: a role for IgG anti-Gal alpha1-3Gal antibodies. *J Immunol* **1996,** *157* (11), 5163-8; (c) Parker, W.; Lin, S. S.; Yu, P. B.; Sood, A.; Nakamura, Y. C.; Song, A.; Everett, M. L.; Platt, J. L., Naturally occurring anti-alpha-galactosyl antibodies: relationship to xenoreactive anti-alpha-galactosyl antibodies. *Glycobiology* **1999,** *9* (9), 865-73.
4. Galili, U.; Shohet, S. B.; Kobrin, E.; Stults, C. L.; Macher, B. A., Man, apes, and Old World monkeys differ from other mammals in the expression of alpha-galactosyl epitopes on nucleated cells. *J Biol Chem* **1988,** *263* (33), 17755-62.
5. Koike, C.; Uddin, M.; Wildman, D. E.; Gray, E. A.; Trucco, M.; Starzl, T. E.; Goodman, M., Functionally important glycosyltransferase gain and loss during catarrhine primate emergence. *Proc Natl Acad Sci U S A* **2007,** *104* (2), 559-64.
6. Peters, B. P.; Goldstein, I. J., The use of fluorescein-conjugated Bandeiraea simplicifolia B4-isolectin as a histochemical reagent for the detection of alpha-D-galactopyranosyl groups. Their occurrence in basement membranes. *Exp Cell Res* **1979,** *120* (2), 321-34.
7. Cairns, T.; Karlsson, E.; Holgersson, J.; Taube, D.; Welsh, K.; Samuelsson, G., Confirmation of a major target epitope of human natural IgG and IgM anti-pig antibodies: terminal galactose alpha -1,3-galactose. *Transplant Proc* **1994,** *26* (3), 1384.
8. Vaughan, H. A.; Dabkowski, P. L.; McKenzie, I. F.; Sandrin, M. S., Biochemical analysis of pig xenoantigens detected by human antibodies. *Transplant Proc* **1993,** *25* (5), 2919-20.
9. (a) Samuelsson, B. E.; Rydberg, L.; Breimer, M. E.; Backer, A.; Gustavsson, M.; Holgersson, J.; Karlsson, E.; Uyterwaal, A. C.; Cairns, T.; Welsh, K., Natural antibodies and human xenotransplantation. *Immunol Rev* **1994,** *141*, 151-68; (b) Bäcker, A. E.; Holgersson, J.; Samuelsson, B. E.; Karlsson, H., Rapid and sensitive GC/MS characterization of glycolipid released Galalpha1,3Gal-terminated oligosaccharides from small organ specimens of a single pig. *Glycobiology* **1998,** *8* (6), 533-545; (c) Hallberg, E. C.; Holgersson, J.; Samuelsson, B. E., Glycosphingolipid expression in pig

aorta: identification of possible target antigens for human natural antibodies. *Glycobiology* **1998,** *8* (7), 637-649.
10. Macher, B. A.; Galili, U., The Galalpha1,3Galbeta1,4GlcNAc-R (alpha-Gal) epitope: a carbohydrate of unique evolution and clinical relevance. *Biochim Biophys Acta* **2008,** *1780* (2), 75-88.
11. Xu, H.; Wan, H.; Zuo, W.; Sun, W.; Owens, R. T.; Harper, J. R.; Ayares, D. L.; McQuillan, D. J., A porcine-derived acellular dermal scaffold that supports soft tissue regeneration: removal of terminal galactose-alpha-(1,3)-galactose and retention of matrix structure. *Tissue Eng Part A* **2009,** *15* (7), 1807-19.
12. Posekany, K. J.; Wiley, J. E.; Gagnon, G. A., A novel method to display [gal alpha1, 3 gal] antigens on human leukemic cells for preparation of anti-leukemia vaccines. *Anticancer Res* **2009,** *29* (6), 2387-92.
13. Mayer, M.; Meyer, B., Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew Chem Int Ed Engl* **1999,** *38* (12), 1784-1788.

Chapter 5. Molecular Recognition of functionalized glycodendrimers by plant lectins: Lectin from Viscum album (Viscumin)

5.1 Introduction

As a general term, we can say that the "valency" of a particle (namely a small molecule, oligosaccharide, protein, nucleic acid, lipid or aggregate of these molecules, a virus, bacterium, or cell) is defined as the number of separated structural units of the same kind that can interact with other particles through ligand–receptor interactions[1].Thus, we can consider that a molecule having a large amount of multiple, tethered and identical copies of binding components (ligands or epitopes) can be classified as a multivalent entity[2]. These multivalent structures (or dendrimers) can be present on a molecular construct (See Figure 5.1) or biological surface that binds in a cooperative way to multiple receptors expressed on a complementary entity.

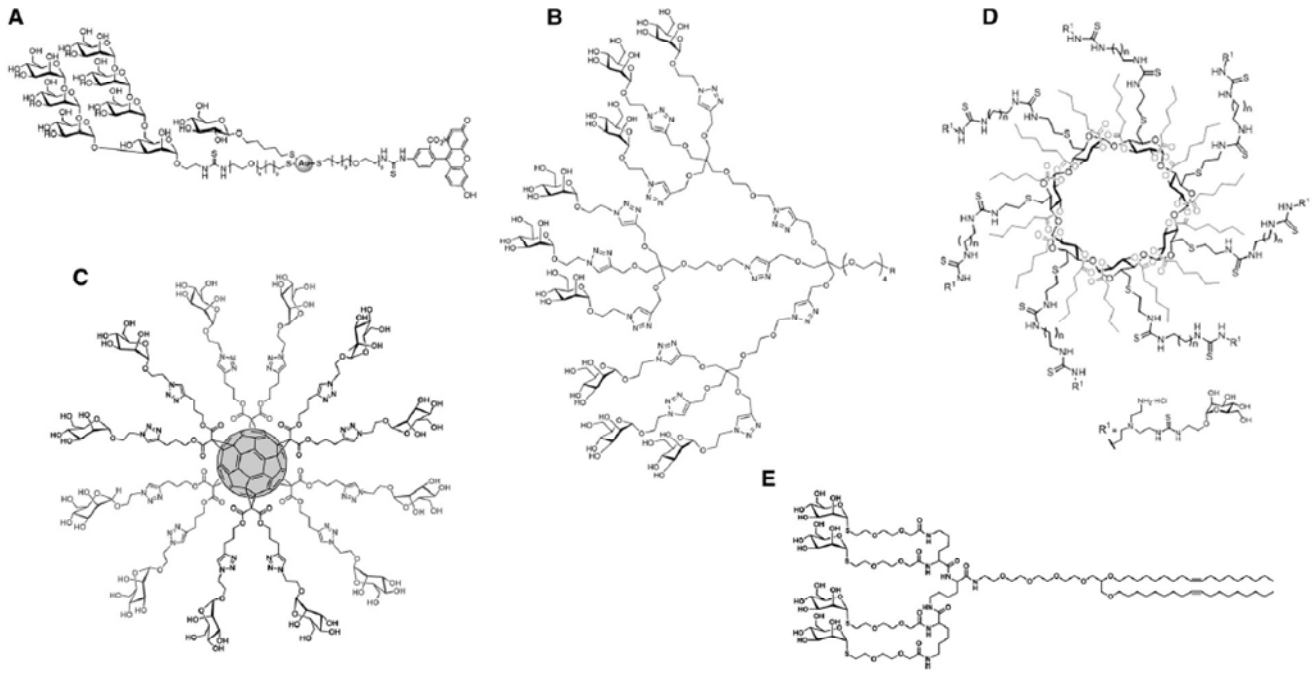


Figure 5.1 (A-E) Examples of different dendrimer scaffold constructions decorated with different carbohydrate structures (mannose residues in this case)[3].

In fact, multivalency in Nature is very often expressed by dendrimeric architectures that represent perhaps the most pervasive topologies observed in vegetal and animal kingdoms[4]. The reasons for this are not entirely clear. For example, trees use fractal dendrimeric patterns above and beneath the ground in order to enhance the

exposure of their leaves to sunlight to harvest light and maximize the photosynthesis process, and of their roots to collect water from the soil. The arterial and central nervous system networks, together with the kidneys and lung structures also consist of a great number of cells growing into dendritic (or dendrimeric) structures in order to gain the largest exchange of material or information with the surrounding tissues[5].

Nowadays, chemical or chemoenzymatic synthesis of multivalent carbohydrate dendrimers is very relevant because carbohydrate structures play critical roles in multiple key cellular events, such as cellular adhesion and recognition, regulation of physiological functions, and pathogenic infections[2]. These carbohydrate-receptor interactions are transformed into very potent attractive forces, dramatically and naturally reinforced, when multiple ligand copies are presented to similarly clustered receptors. This cooperative effect is called "glycocluster or dendritic effect" that may offer numerous benefits in terms of affinity and receptor selectivity versus monovalent interaction. The presentation of the sugar epitopes as multiple copies on an appropriate scaffold (molecular, dendritic, polymeric) creates a multivalent display that can efficiently mimic the natural mode of affinity enhancement that arises from multiple interactions between the binding proteins and the carbohydrate ligands. The ligand-binding specificity of these complexes with lectins is determined by glycan microarrays[6] and validated by a range of binding assays such as ELISA (enzyme-linked immunosorbent assay), ITC(isothermal titration calorimetry), SPR (surface plasmon resonance), X-ray crystallography, NMR spectroscopy and analogous techniques[2].

Glycodendrimers have a wide variety of known applications. These glyco-scaffolds are useful tools in drug discovery has made it possible to participate in extracellular and intracellular biochemical processes. Their applications in biological systems are fully studied; as for example to used them against viral infections[7], against bacterial and plant toxins[8] and bacterium adhesion to human cells[8a, 9] (See Figure 5.2), and for immune system stimulation and vaccine approach[3].

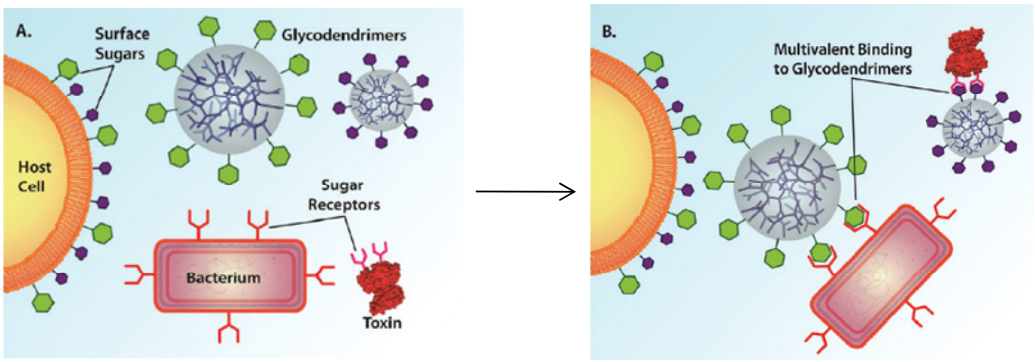


Figure 5.2 Example of a multivalent binding event with a bacterium or a bacterial toxin and a glycodendrimer. (A) Depiction of a host eukaryotic cell with surface green and purple sugar groups, two glycodendrimers with sugar groups in their surface, a bacterial toxin with receptors for the purple sugar groups, and a generic bacterium with receptors for the green sugar groups. (B) Representation of the binding of the bacterium and the toxin to the glycodendrimer through multivalent interactions to prevent infection of the host cell[8a].

5.2 Objectives

The high-resolution features of the recognition process of synthetic lactose- and 6-sialyllactose glycodendrimers by viscumin, a toxic plant lectin (VAA) have been elucidated. The epitope of the ligand has been dissected as well as the possible multivalent capability of the different glycodendrimers to interact with more than one unit of the lectin. In principle, this fact should increase the binding affinity as well as establish the possibility to generate larger networked complexes.

5.3 The bound state

5.3.1 The recognition of lactose-functionalized glycodendrimers by Viscumin; A combined STD-docking approach

STD experiments

Figure 5.3 shows the STD analysis for ligand DA2 in the presence of VAA, together with the assignment of the key signals. Clear STD signals were observed, as depicted in the Figure. The strongest STD effect was detected for the sugar protons for the Gal residue (with a STD absolute intensity average of approximately 9%; see Figure

5.3B), strongly suggesting that this region of the ligand is in closer contact with the protein. The protons belonging to the Glc residue displayed less intense STD signals, around 4%, as identified in Fig. 5.3B.

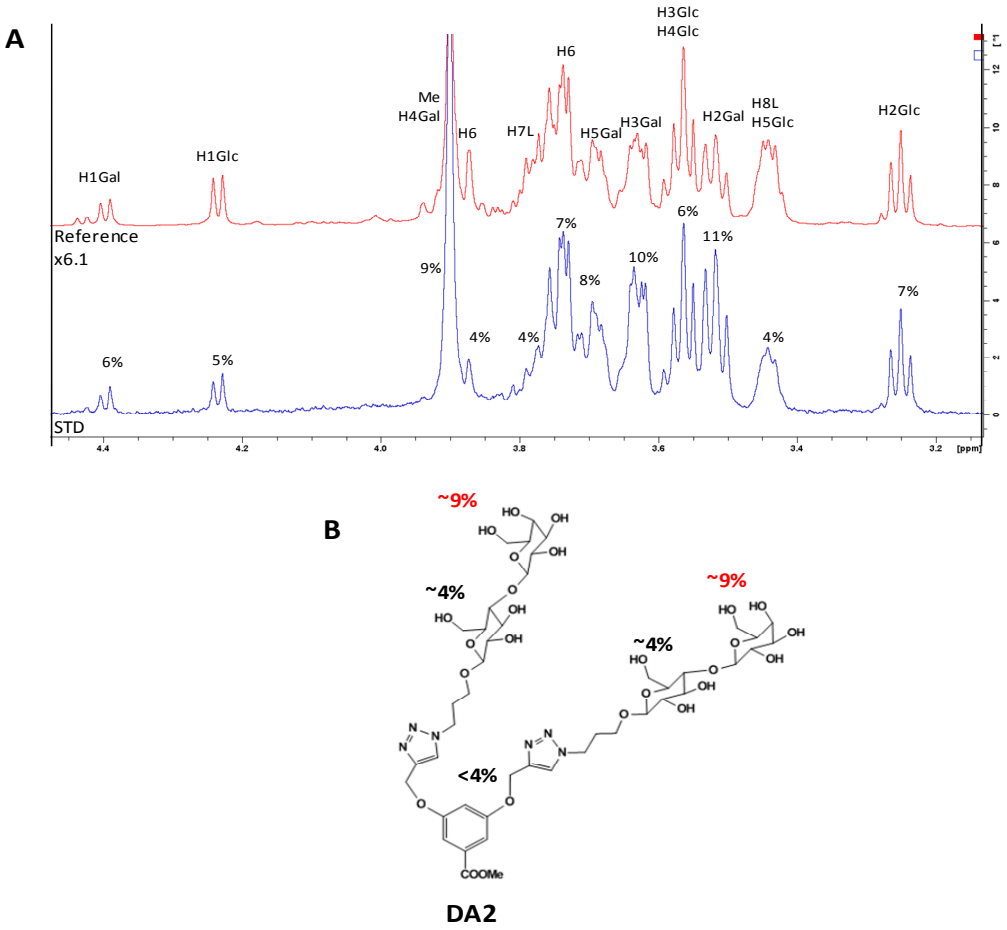


Figure 5.3 (A) STD spectra (bottom) and off-resonance or reference spectra (top) of the sample that contains DA2 glycodendrimer in the presence of viscumin, with a 50:1 molar ratio. The absolute STD percentages are given for the corresponding proton signals. (B) Schematic structure of DA2 along with the average values of the absolute STD intensities.

Figure 5.4 shows the STD results for DA3 ligand in the presence of VAA, together with the assignment of the key chemical shifts. Clear STD signals were observed, and the strongest STD effect arose again from the sugar protons for the Gal residue with a STD absolute intensity average of approximately 8% (see Figure 5.4B). In fact, the rest of protons from the dendrimer displayed significantly weaker STDs, around 3%, as depicted in Fig. 5.4B.

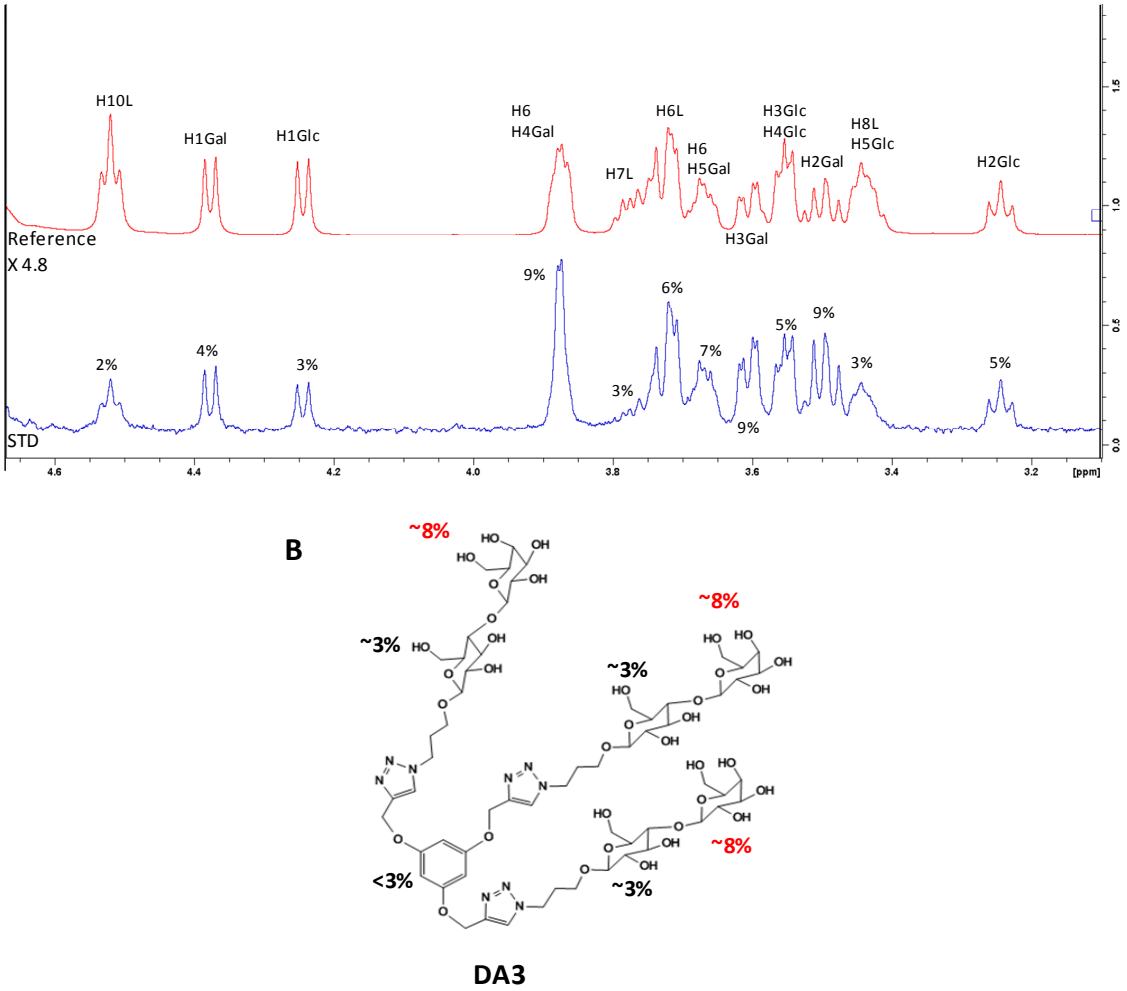


Figure 5.4 (A) STD spectra (bottom) and off-resonance or reference spectra (top) of the sample that contains DA3 in the presence of viscumin, with a 50:1 molar ratio. The absolute STD percentages are given for the corresponding proton signals. (B) Schematic structure of DA3 along with the average values of the absolute STD intensities.

Figure 5.5 displays the STD analysis for DA4 ligand, using the same protocol. Similar to the observations for DA2 and DA3 glycodendrimers, again the strongest STD effect arose from the Gal moiety with STD absolute intensity averages of approximately 5% (Figure 5.5B). Again, the Glc residue and the aromatic linker protons showed STD, although in lesser extent than for the Gal residue (STD absolute intensity average of approximately 3 and 4% respectively; see Figure 5.5B).

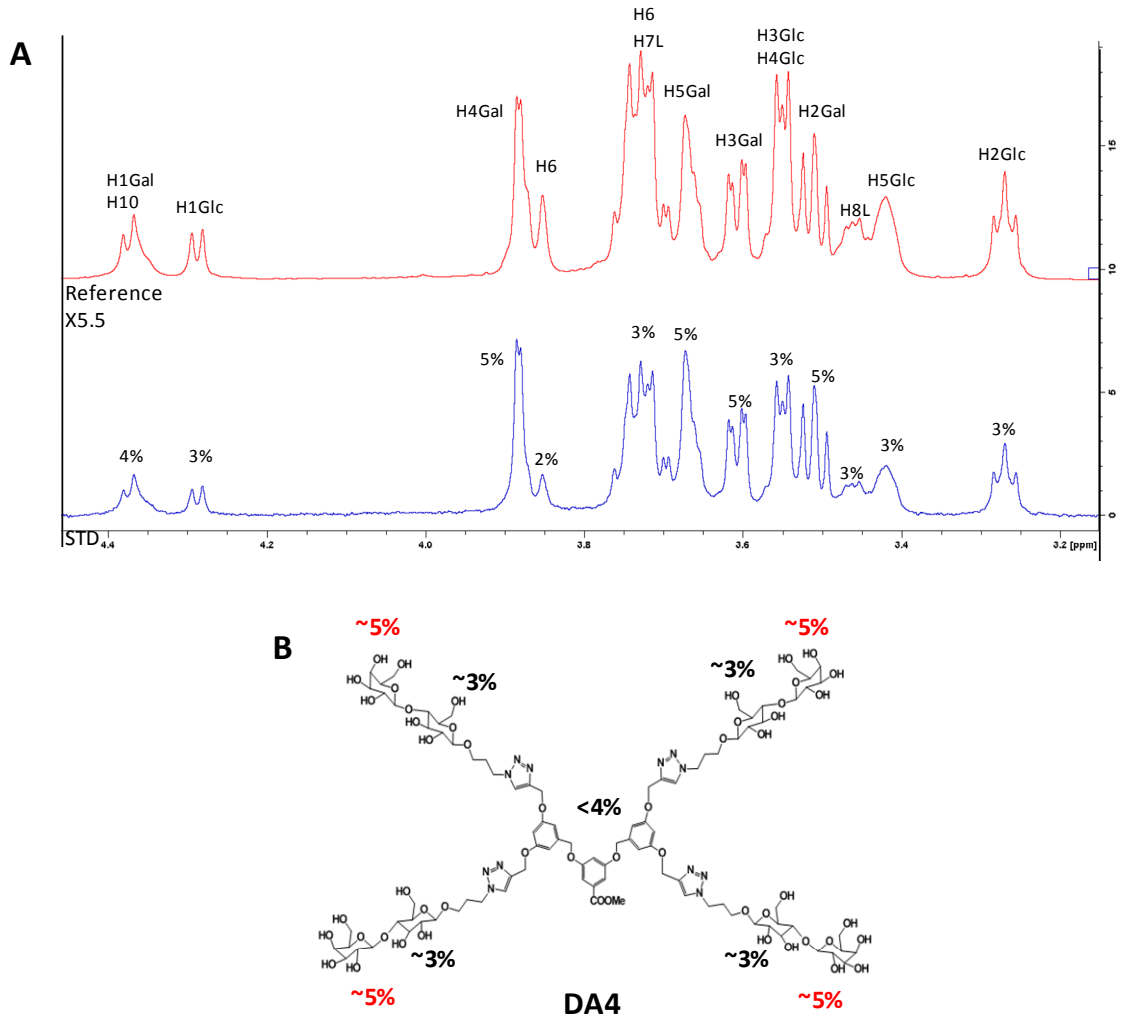


Figure 5.5 (A) STD spectra (bottom) and off-resonance or reference spectra (top) of the sample that contains DA4 in the presence of viscumin with a molar ratio50:1. The absolute STD percentages are given for the corresponding proton signals. (B) Schematic structure of DA4 along with the average values of the absolute STD intensities.

Docking model

These experimental observations were validated using a three-dimensional model of the complex, which was built using the method described in the experimental section. The well known VAA binding site (the TYR binding site[10]) was employed for the docking protocol. DA2, DA3 and DA4 ligands basically produced similar results in the presence of viscumin. Figure 5.6 displays the aminoacids that are implicated in the recognition of the glycodendrimer. Most of the key interactions involve the Gal moiety, as deduced from the STD NMR analysis. It should be pointed out the existence of a bidentate hydrogen bond between ASP235 with OH3 and OH4 of the Gal residue, as

well as a second hydrogen bond between ASN256 and the OH3 of the Gal moiety. Additionally, a clear CH-π stacking interaction between TYR249 and the Gal residue takes place. Interestingly, GLN238 establishes a hydrogen bond with OH3 of Glc residue.

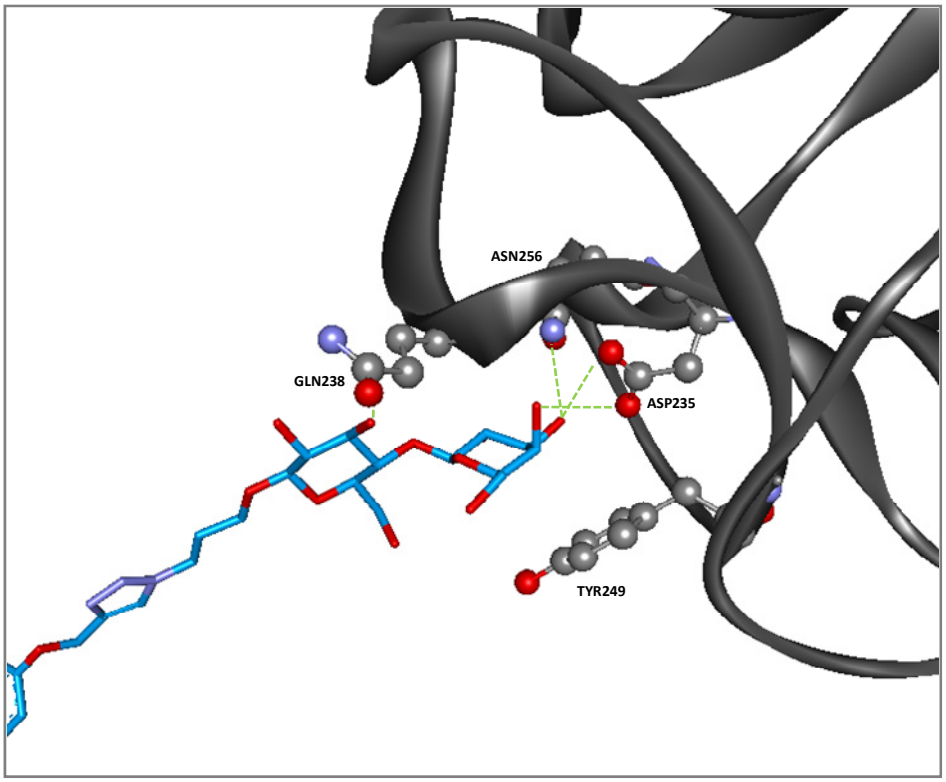


Figure 5.6 Docking model of the region (lactose) of the glycodendrimer DA2 that binds to the viscumin TYR binding site. The aminoacids providing better contacts with the ligand (GLN238, ASN256, ASP235, TYR249) are highlighted.

The possibility of simultaneous binding of two, three, or four VAA moieties to the ligand was also explored using the docking protocol. The results are given in Figures 5.7, 5.8 and 5.9 respectively. The built model suggests the possibility of simultaneous binding of two and three VAA units in the case of DA2 and DA3 glycodendrimers respectively (see Figure 5.7 and 5.8).However, only of a maximum of three VAA units can be bound to DA4 ligand. Indeed, four simultaneous units would rise to important steric conflicts (See Image B, Figure 5.9), thus precluding the existence of larger multivalent complexes. Nevertheless, under the experimental conditions employed for the NMR experiments, with excess of ligand over the lectin entity, it is evident that the possibility of formation of high-order complexes is highly unexpected. Very probably, the NMR data reflect the presence of 1:1 complexes in fast dissociation in the relaxation time scale (as required for the presence of STD).

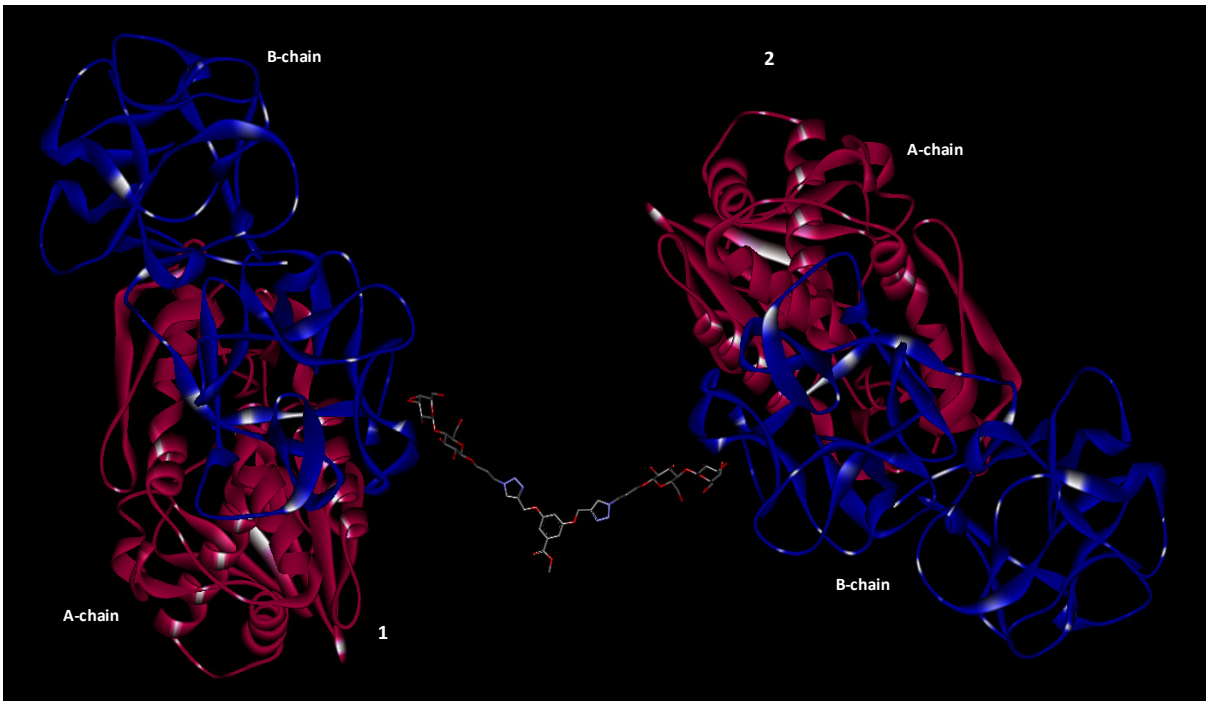


Figure 5.7 Docking model of the glycodendrimer DA2bound to two units of viscumin.

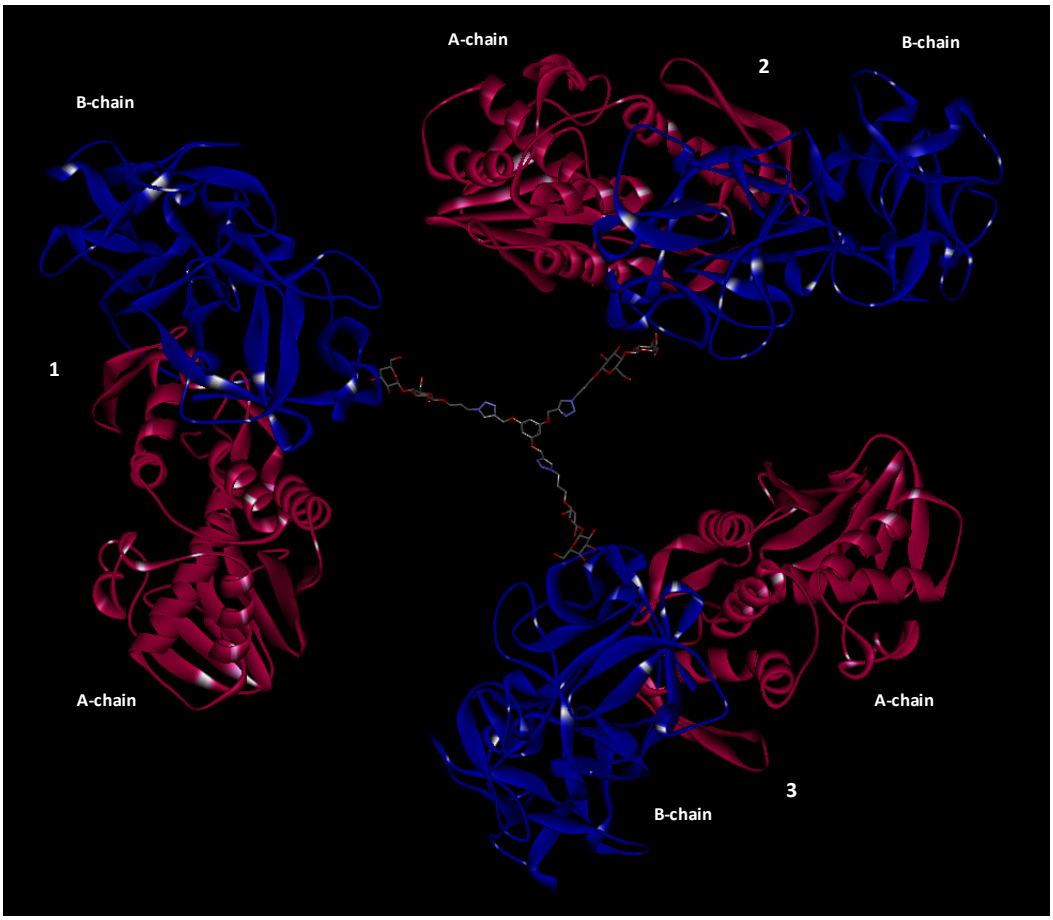


Figure 5.8 Docking model of the glycodendrimer DA3 bound to three units of viscumin.

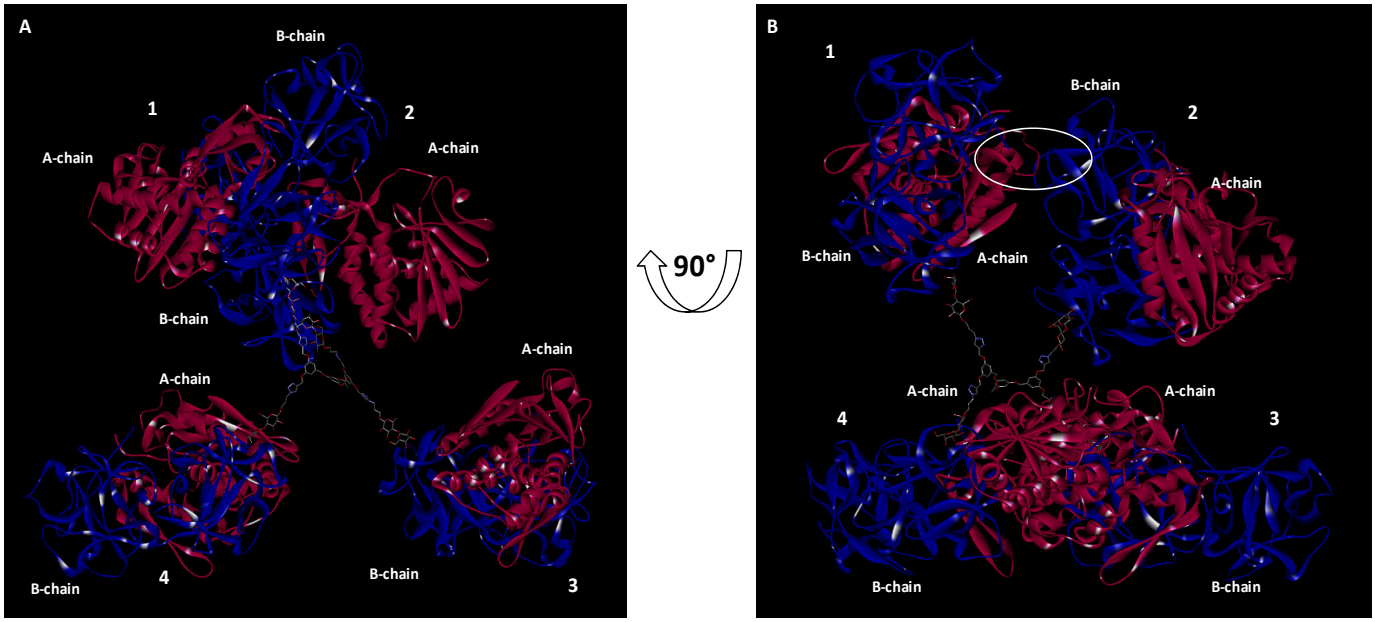


Figure 5.9 (A) Docking model of the glycodendrimer DA4 bound to four units of viscumin. (B) Different view of the docking model of the glycodendrimer DA4 bound to four units of viscumin.

5.3.2 The recognition of 6'sialyllactose-functionalized glycodendrimers by Viscumin; A combined STD-docking approach

STD experiments

Figure 5.10 shows the STD results in the case of ligand DA3-2,6-Ac in the presence of Viscumin (VAA), together with the assignment of the key chemical shifts. Clear STD signals were again observed, and the strongest STD effect was detected for the sugar protons for Gal residue followed by Glc and sialic acid residues (with a STD absolute intensity average of approximately 3%; see Figure 5.10C), strongly suggesting that this region of the ligand is in closer contact with the protein. However, also several protons from the glucose and sialic acid residues displayed very clear STDs, around 2%, as identified in Fig. 5.10C. The aromatic linker also received saturation of the protein, less than 2%.

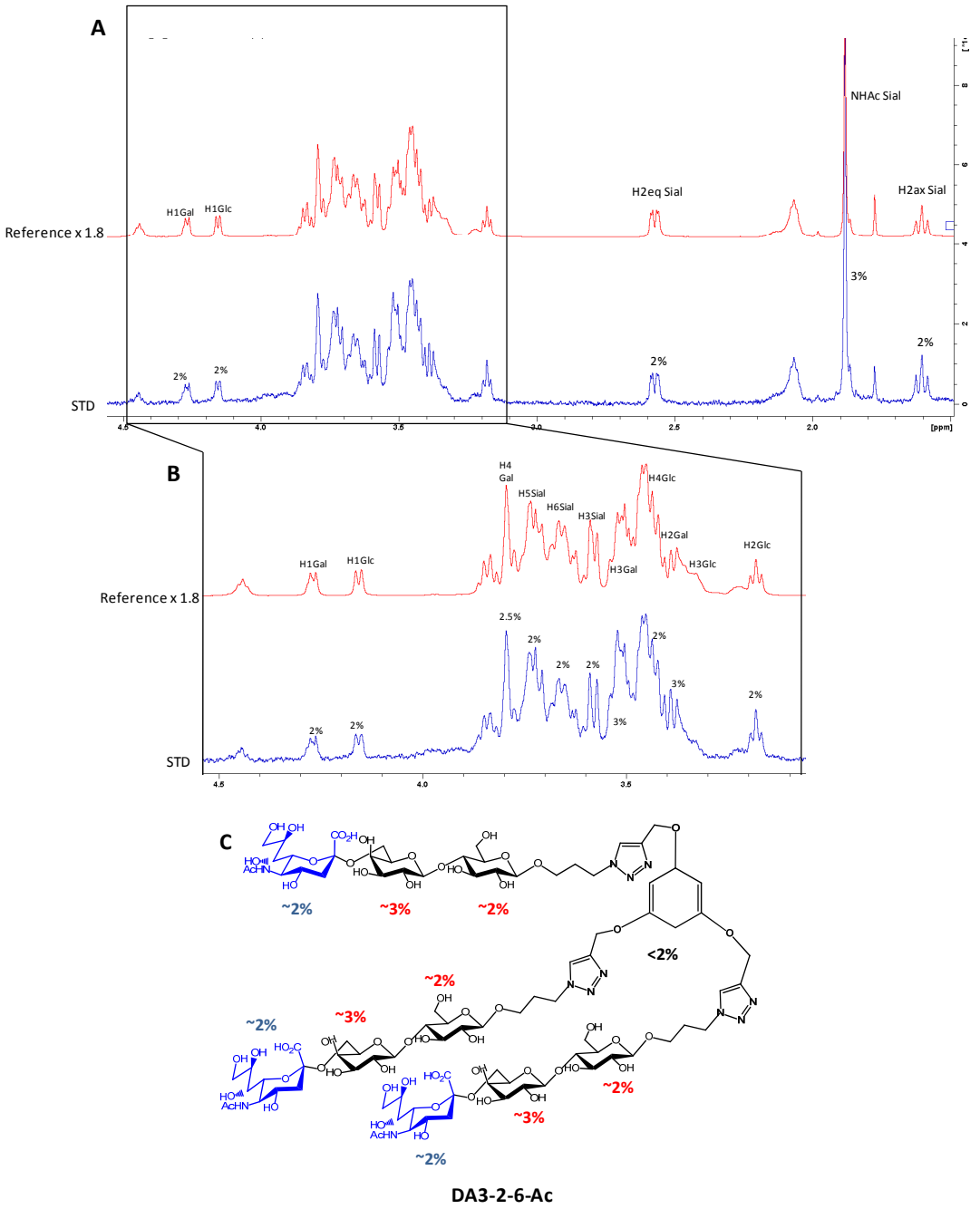


Figure 5.10 (A) STD spectra (bottom) and off-resonance or reference spectra (top) of the sample that contains DA3-2,6-Ac in the presence of viscumin with a molar ratio 100:1. The absolute STD percentages are given for the corresponding proton signals. (B) Extension of the proton sugar region with the corresponding STD absolute values. (C) Schematic structure of DA3-2,6-Ac along with the average values of the absolute STD intensities.

Figure 5.11 shows the STD results for ligand DA4-2,6-Ac in the presence of VAA, together with the assignment of the key signals. Clear STD signals were observed, which are also depicted in the Figure. The strongest STD effect was detected for the sugar protons for both Glc and Gal residues (with a STD absolute intensity average of approximately 5%; see Figure 5.11C), strongly suggesting that this region of

the ligand is in closer contact with the protein. However, also several protons from the sialic acid residue displayed very clear STDs, around 3%, as identified in Fig. 5.11C. The aromatic linker also received saturation of the protein.

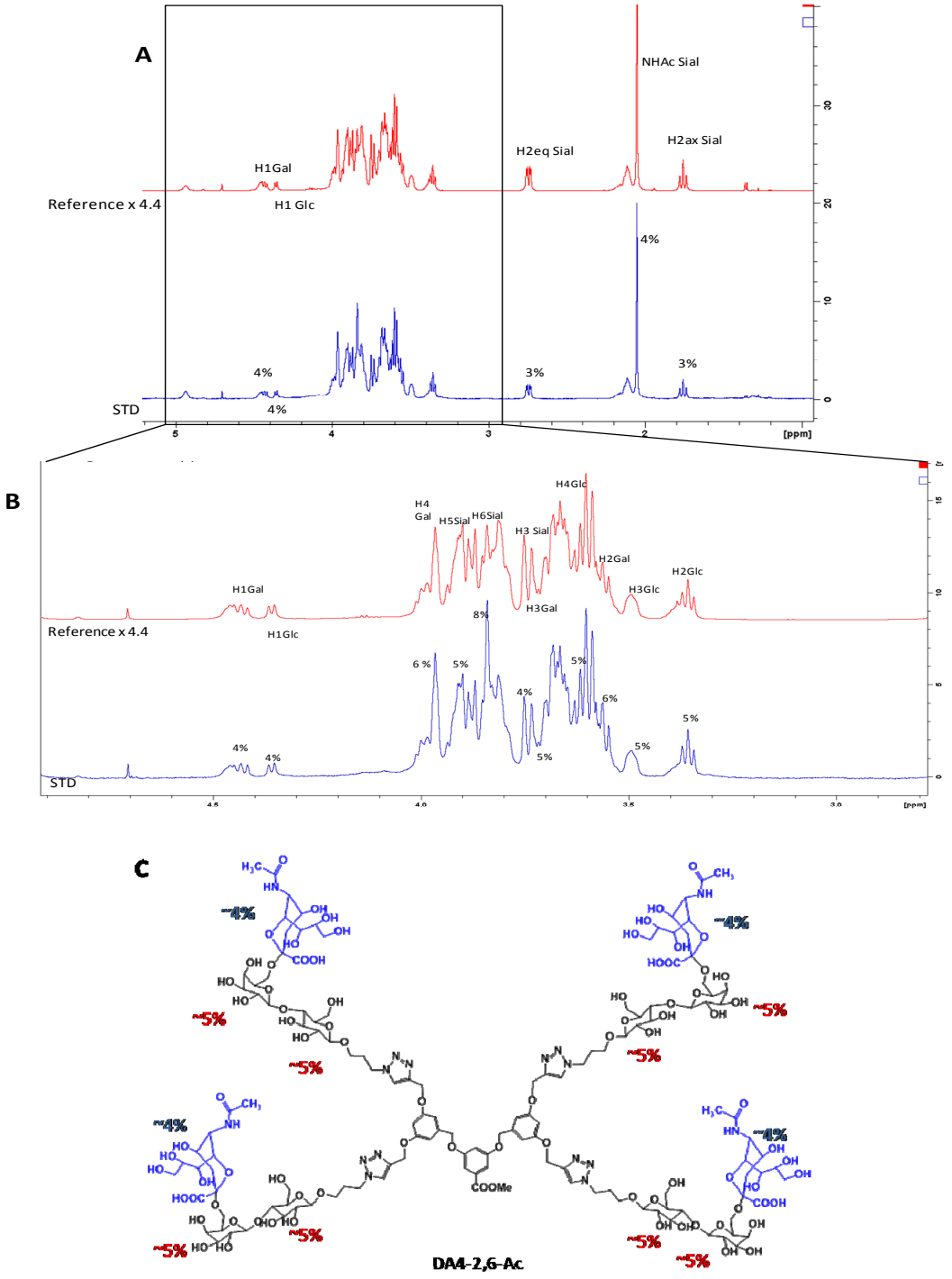


Figure 5.11 (A) STD spectra (bottom) and off-resonance or reference spectra (top) of the sample that contains DA4-2,6-Ac in the presence of viscumin with a molar ratio 100:1. The absolute STD percentages are given for the corresponding proton signals. (B) Extension of the proton sugar region with the corresponding STD absolute values. (C) Schematic structure of DA4-2,6-Ac along with the average values of the absolute STD intensities.

Figure 5.12 displays the STD analysis for ligand DA4-2,6-Gc, using the same protocol. Similar to the observations for DA4-2,6-Ac, the strongest STD effect arose from the sialyllactose moiety with Glc, Gal and Sial STD absolute intensity averages of approximately 3, 3 and 4% respectively (Figure 5.12C). Again, the aromatic linker protons showed STD, although in lesser extent than for the DA4-2,6-Ac analogue.

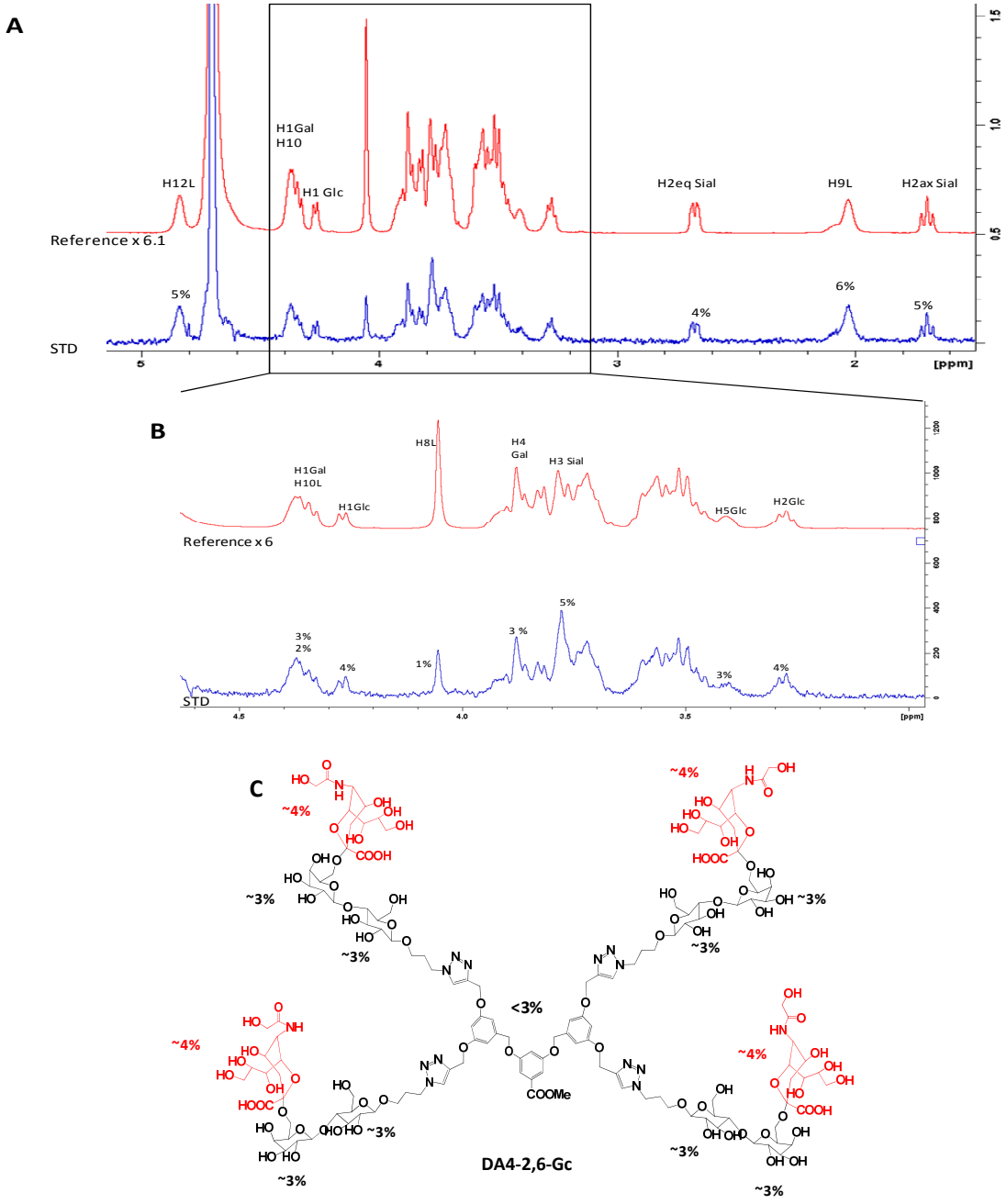


Figure 5.12. (A) STD spectra (bottom) and off-resonance or reference spectra (top) of the sample that contains DA4-2,6-Gc in the presence of viscumin with a molar ratio 100:1. The absolute STD percentages are given for the corresponding proton signals. (B) Extension of the proton sugar region with the corresponding STD absolute values. (C) Schematic structure of DA4-2,6-Gc along with the average values of the absolute STD intensities.

Docking model

These experimental observations were also validated using a three-dimensional model of the complex, which was built using the method described above. The different ligands basically produced similar results in the presence of viscumin. Figure 5.13 displays the aminoacids that are implicated in the recognition of the glycodendrimer. Interestingly, SER200 and GLN238 establish hydrogen bonds with OH3 and the carboxylic acid of the sialic acid residue, respectively. It should be pointed out the existence of a bidentate hydrogen bond between ASP235 with OH3 and OH4 of the Gal residue, as well as a second hydrogen bond between ASN256 and the OH3 of the Gal moiety. Additionally, a clear CH-π stacking interaction between TYR249 and the Gal residue takes place.

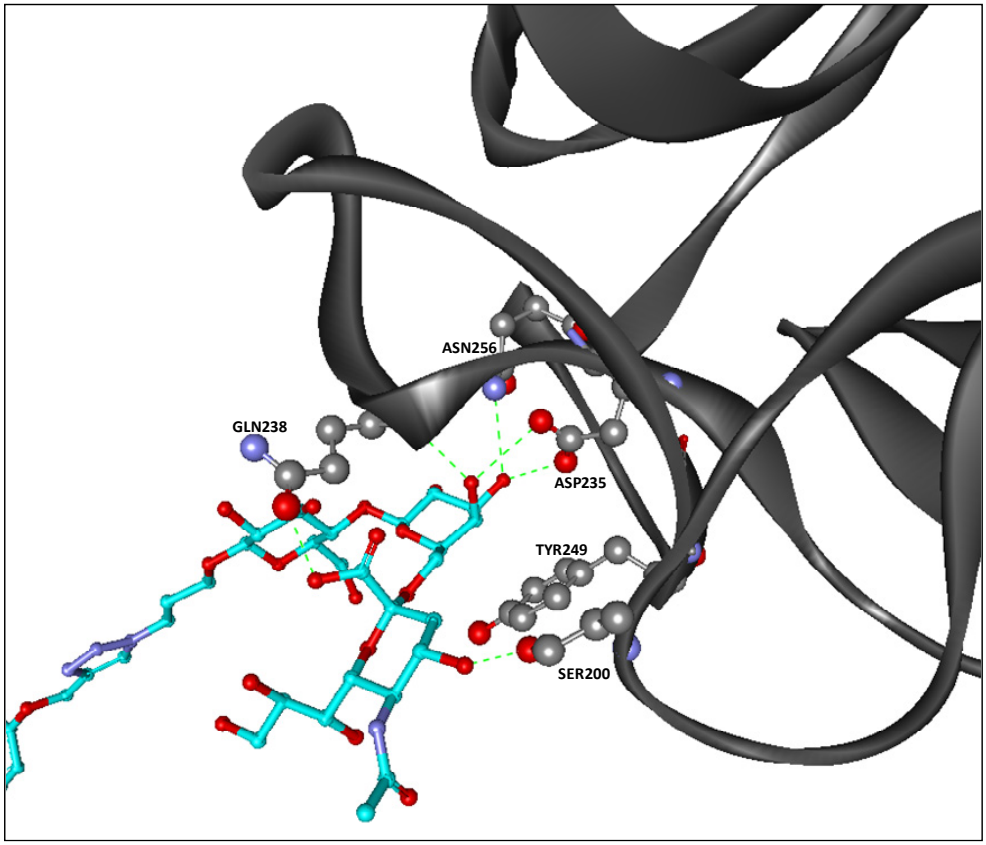


Figure 5.13 Docking model of the region (sialyllactose) of the glycodendrimer DA4-2,6-Ac that binds to the viscumin TYR binding site. The aminoacids providing better contacts with the ligand (GLN238, ASN256, ASP235, TYR249, SER200) are highlighted.

The possibility of simultaneous binding of two, three, or four VAA moieties to the ligand was also explored using the docking protocol. The results are given in Figures 5.14 and 5.15. The built model suggests the possibility of simultaneous binding of three VAA units in DA3-2,6-Ac, but only of a maximum of two VAA units in the case of DA4 ligands. Only with DA4 glycodendrimers, the simultaneous binding of three lectins would give rise to important steric conflicts, thus precluding the existence of larger multivalent complexes.

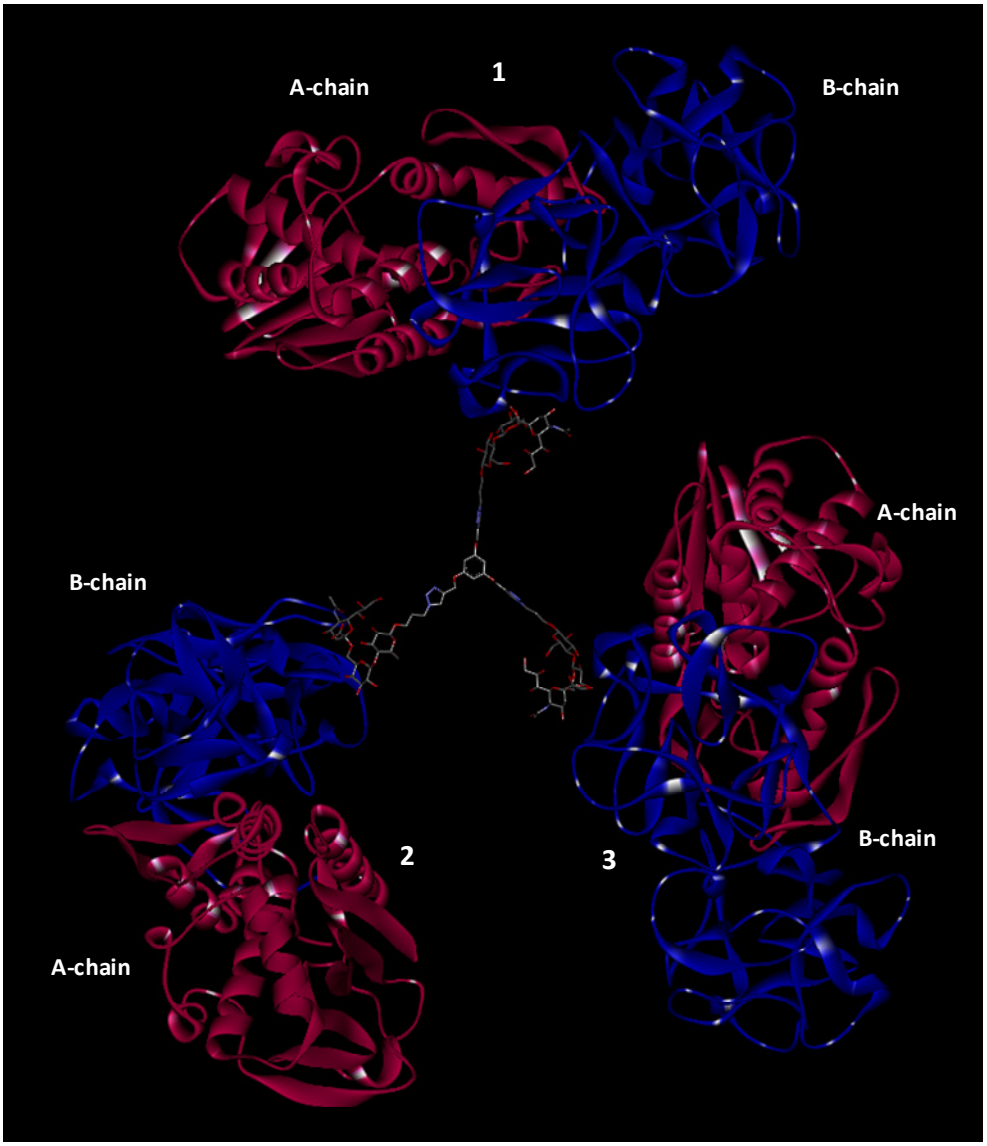


Figure 5.14 Docking model of the glycodendrimer DA3-2,6-Ac bound to four units of viscumin (Top) and the docking models of the corresponding non precluded complexes (Down).

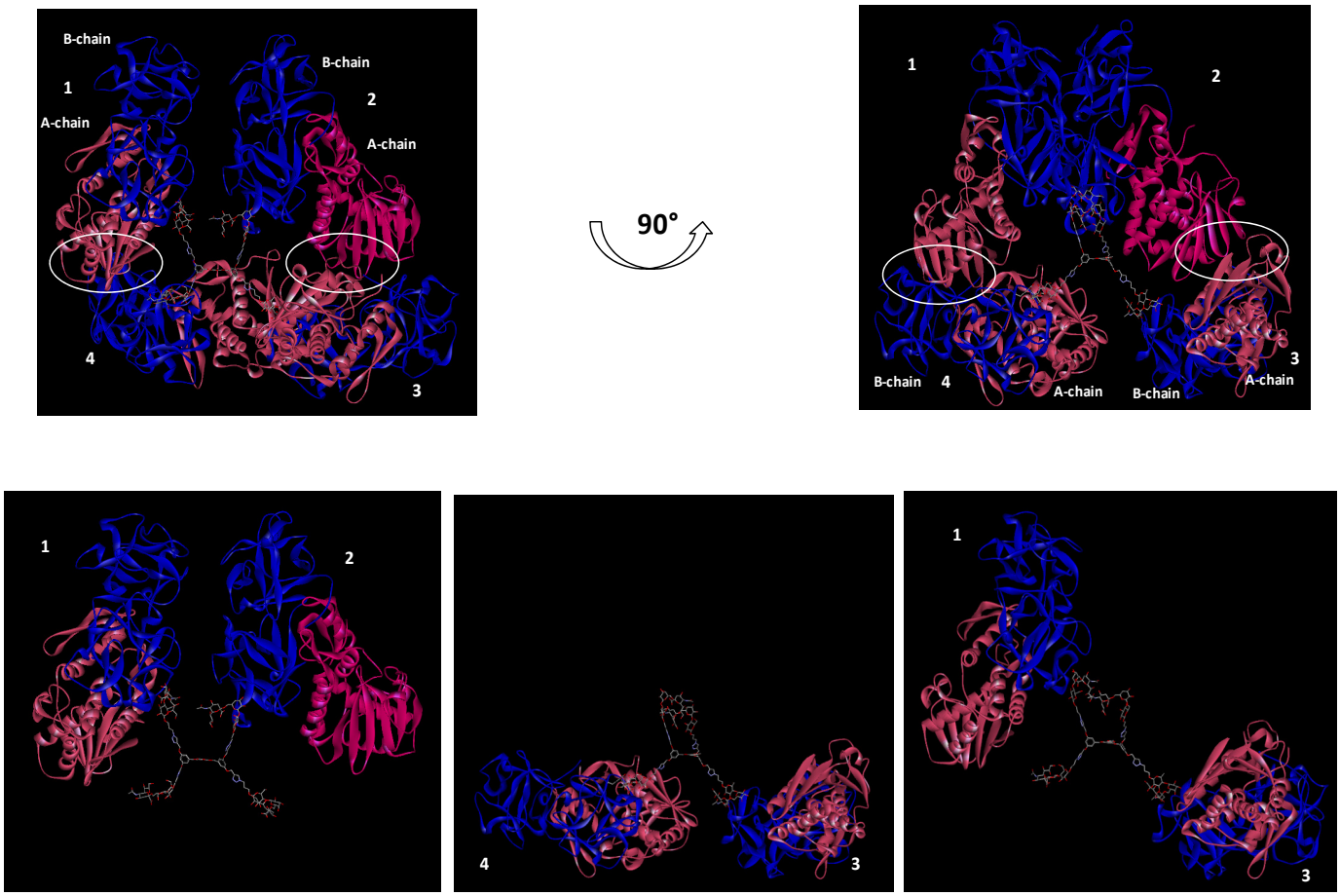


Figure 5.15 Docking model of the glycodendrimer DA4-2,6-Ac bound to four units of viscumin (Top) and the docking models of the corresponding non precluded complexes (Down).

On the other hand, no experimental evidences of simultaneous binding could be extracted from the NMR experiments in a non-ambiguous manner, using different VAA/ligand molar ratios in DOSY experiments (see Figure 5.16). Observation of the diffusion coefficient of the NMR signals of DA4-2,6-Ac ligand in the presence of different amounts of VAA showed a significant increase in their corresponding diffusion coefficient. However, given the excess of the ligand and the linear averaging of the diffusion coefficient, these observations could be due to the presence of either 1:1 or 1:2 complexes between the ligand and VAA.

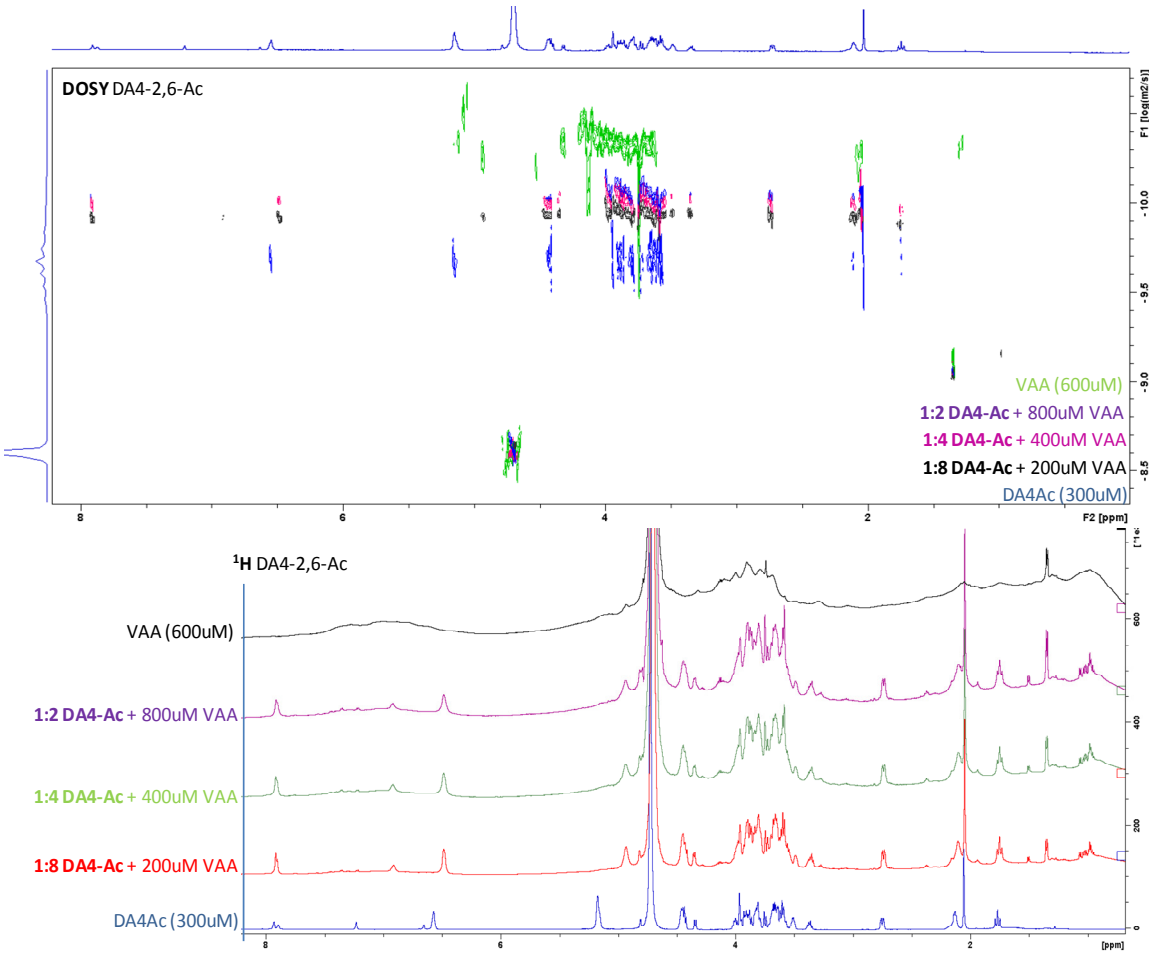


Figure 5.16 (Top) DOSY experiments of glycodendrimer DA4-2,6-Ac in the presence of Viscumin (DA4-2,6-Ac:VAA ratios 1:8, 1:4 and 1:2). (Down) **^{1}H**-NMR spectra of samples described on top.

5.4 Conclusions

Therefore, the reported lactose- and 6'sialyllactose-functionalized glycodendrimers can be easily accommodated by this lectin, providing enhanced binding. NMR experiments assisted by molecular modeling permitted the identification of the key sugar-lectin binding features, while the modeling protocol suggested a model for the possibility of multivalent interactions of these ligands with this lectin.

5.5 Materials and Methods

Methods

NMR experiments

The NMR study of compounds DA2, DA3, and DA4in water solution was performed using 1mM samples in D_2O. Standard 2D-methods, such as TOCSY (with mixing times of 30 and 80ms), ROESY (300 ms), NOESY (400 and 600 ms), and HSQC experiments were employed to assign the NMR chemical shifts, using Bruker Avance 500 and 600 MHz (equipped with a triple channel cryoprobe) NMR spectrometers.

Viscumin (Viscum Album Agglutinin, VAA) samples for saturation-transfer difference (STD) experiments were prepared in phosphate-buffered saline in deuterated water ($NaH_2PO_4$10mM, NaCl137mM andKCl2.7mM; pH7.3 for Viscumin, uncorrected) using ligand/lectin ratios of 50:1. The temperature was 298 K. Data from one-dimensional (1D) STD experiments were acquired at 600 and 500 MHz using a Gaussian pulse (49 ms) cascade separated by 1 ms delays. In all cases, the on-resonance frequency was set at the aliphatic region (-0.2ppm), and the experiments were recorded with a saturation time of 2 s. The off-resonance frequency was always set at 100 ppm.

The NMR study of DA3-2,6-Ac, DA4-2,6-Ac and DA4-2,6-Gc samples in water solution was performed using 1mM samples in D_2O. Standard 2D-methods, such as TOCSY (with mixing times of 30 and 80ms), ROESY (300 ms), NOESY (400 and 600 ms), and HSQC experiments were employed to assign the NMR chemical shifts, using Bruker Avance 500 and 600 MHz (equipped with a triple channel cryoprobe) NMR spectrometers.

Viscumin (Viscum Album Agglutinin, VAA) samples for saturation-transfer difference (STD) experiments were prepared in deuterated water (NaH_2PO_4 10mM, NaCl137mM andKCl2.7mM; pH7.3 for Viscumin, uncorrected) using ligand/lectin ratios of 100:1. The temperature was 298 K. Data from one-dimensional (1D) STD experiments were acquired at 600 and 500 MHz using a Gaussian pulse (49 ms) cascade separated by 1 ms delays. In all cases, the on-resonance frequency was set at the aliphatic regions (0.5 ppm), and the experiments were recorded with a saturation time of 2 s. The off-resonance frequency was always set at 100 ppm.

Molecular Modelling

The templates for the modeling procedures were built based on the reported X-ray structure of Viscumin complexed with lactose (pdb code: 1PUU). The structures of the glycodendrimers were superimposed in the TYR249 binding site,using the most populated conformation found for the free state (according to a standard NOE/molecular modeling approach). The complex structure was then submitted to a short molecular dynamics (MD) run, followed by energy minimization with a low gradient convergence threshold (0.05) in 5000 steps. In all cases, the OPL2005 force field[11] was employed, as integrated in the Schroedinger MAESTRO suite of programms[12].The protocol has been already described[12].

5.6 References

1. Mammen, M.; Choi, S.-K.; Whitesides, G. M., Polyvalent interactions in biological systems: Implications for design and use of multivalent ligands and inhibitors. *Angew Chem Int Ed Engl* **1998,** *37*, 2754-2794.
2. Chabre, Y. M.; Roy, R., DESIGN AND CREATIVITY IN SYNTHESIS OF MULTIVALENT NEOGLYCOCONJUGATES. *Adv Carbohydr Chem Biochem* **2010,** *63*, 165-393.
3. Lepenies, B.; Lee, J.; Sonkaria, S., Targeting C-type lectin receptors with multivalent carbohydrate ligands. *Adv Drug Deliv Rev* **2013,** *65* (9), 1271-81.
4. Thompson, D., On growth and form. *On growth and form, Cambridge University Press, London* **1987**.
5. Ganong, W. F., Review of medical physiology. *Review of medical physiology, 15th edn. Prentice-Hall, New York* **1991**.
6. Parera Pera, N.; Branderhorst, H. M.; Kooij, R.; Maierhofer, C.; van der Kaaden, M.; Liskamp, R. M.; Wittmann, V.; Ruijtenbeek, R.; Pieters, R. J., Rapid screening of lectins for multivalency effects with a glycodendrimer microarray. *Chembiochem* **2010,** *11* (13), 1896-904.
7. (a) Sanchez-Navarro, M.; Rojo, J., Targeting DC-SIGN with carbohydrate multivalent systems. *Drug News Perspect* **2010,** *23* (9), 557-72; (b) Luczkowiak, J.; Sattin, S.; Sutkeviciute, I.; Reina, J. J.; Sanchez-Navarro, M.; Thepaut, M.; Martinez-Prats, L.; Daghetti, A.; Fieschi, F.; Delgado, R.; Bernardi, A.; Rojo, J., Pseudosaccharide functionalized dendrimers as potent inhibitors of DC-SIGN dependent Ebola pseudotyped viral infection. *Bioconjug Chem* **2011,** *22* (7), 1354-65.
8. (a) Mintzer, M. A.; Dane, E. L.; O'Toole, G. A.; Grinstaff, M. W., Exploiting dendrimer multivalency to combat emerging and re-emerging infectious diseases. *Mol Pharm* **2012,** *9* (3), 342-54; (b) Munoz, F. J.; Santos, J. I.; Arda, A.; Andre, S.; Gabius, H. J.; Sinisterra, J. V.; Jimenez-Barbero, J.; Hernaiz, M. J., Binding studies of adhesion/growth-regulatory galectins with glycoconjugates monitored by surface plasmon resonance and NMR spectroscopy. *Org Biomol Chem* **2010,** *8* (13), 2986-92.
9. Bernardi, A.; Jimenez-Barbero, J.; Casnati, A.; De Castro, C.; Darbre, T.; Fieschi, F.; Finne, J.; Funken, H.; Jaeger, K. E.; Lahmann, M.; Lindhorst, T. K.; Marradi, M.; Messner, P.; Molinaro, A.; Murphy, P. V.; Nativi, C.; Oscarson, S.; Penades, S.; Peri, F.; Pieters, R. J.; Renaudet, O.; Reymond, J. L.; Richichi, B.; Rojo, J.; Sansone, F.; Schaffer, C.; Turnbull, W. B.; Velasco-Torrijos, T.; Vidal, S.; Vincent, S.; Wennekes, T.; Zuilhof, H.; Imberty, A., Multivalent glycoconjugates as anti-pathogenic agents. *Chem Soc Rev* **2013,** *42* (11), 4709-27.
10. (a) Kaptein, R.; Dijkstra, K.; Nicolay, K., Laser photo-CIDNP as a surface probe for proteins in solution. *Nature* **1978,** *274* (5668), 293-4; (b) Jimenez, M.; Andre, S.; Siebert, H. C.; Gabius, H. J.; Solis, D., AB-type lectin (toxin/agglutinin) from mistletoe: differences in affinity of the two galactoside-binding Trp/Tyr-sites and regulation of their functionality by monomer/dimer equilibrium. *Glycobiology* **2006,** *16* (10), 926-37.
11. Kaminski, G.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L., Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001,** *105*, 6474-6487.
12. Maestro, A Powerful, All-Purpose Molecular Modeling Environment, Version 8.5. *Schroedinger, LLC* **2008,** *New York*.

Chapter 6. Conclusions

1. We have demonstrated that the combination of STD and trNOESY NMR and MD simulations provide important information about the molecular recognition of synthetic oligosaccharides towards two specific IgG C7-37 and IgM G19-2 antibodies. Our study has provided experimental evidences on the individual importance of the different residues, depending on its position at the oligosaccharide sequence, also for longer oligosaccharides, and their relevance for the molecular recognition process. The information derived from STD-NMR experiments clearly showed evidences of the direct involvement of residue C followed by its neighbor residues (Rha*p*B-**Rha*p*C**-GlcNAc*p*D) in every synthetic oligosaccharide in the presence of both mAbs.

2. trNOESY-NMR experiments provided structural information about the major Φ/Ψ values of each glycosidic linkage and therefore, the existing conformational equilibrium in the bound state, which was compared to that existing in the free state. The data indicate that the molecular recognition events involve conformational selection processes.

3. The analysis of the computed molecular surfaces showed that the major polar zones correspond to the non-interacting region of the ligand. Analysis of molecular surfaces showed that the major non-polar areas of the ligand are close to the antibody surface.

4. The preparation of *N*-glycans facilitates definition of the docking sites of lectins on complex glycans. The nature of the target epitope strictly depends on the nature of the tested lectin, which can select distinct determinants of the complex-type saccharide chain with exquisite specificity, depending on the relative presentation of the different residues and the architecture of the binding site.

5. We have underscored the importance of the coexistence of binding sites from different lectin domains for achieving glycan binding by comparing a single hevein domain and WGA, a tetramer. The monomeric unit does not recognize the large glycan, with the branching at the βMan residue being the key structural element for precluding the interaction. In contrast, the multidomain lectin WGA binds the terminal reducing end of either non- or sialylated-*N*-glycan owing to

the possibility of achieving additional inter-site stabilizing interactions from one of the neighboring hevein domains. Thus, in naturally occurring *N*-glycopeptides, binding occurs via the *N*-glycan–Asn stem region. As demonstrated, on the level of small glycans, WGA can accommodate sialic acid residues. Indeed, using α2,6-sialyllactose, we clearly demonstrate that this small trisaccharide is recognized by WGA, but the situation is different at the level of *N*-glycans.

6. Work with VAA has yielded insights into how α2,6-sialylation can be accommodated by this lectin, providing enhanced binding in natural *N*-glycans when compared to the non-sialylated LacNAc-terminated analogue. Our methodology has permitted identification of the direct involvement of the sialic acid, as well as a network of interactions.

7. Conversely, the results for the MAL–MAH mixture have permitted definition of the role of the α2,6-linked sialic acid residue in blocking binding of the corresponding sialylated-*N*-glycan. The binding to the LacNAc terminus of the non-sialylated-*N*-glycan is detectable, providing the control reference for positive binding.

8. Four oligosaccharides related to the glycotope antigen have been shown to interact with different scFv glycotope-binding antibody fragments. Using STD experiments, we have shown that the presence of 3-O-α-*DGal* A' moiety is a decisive feature for the binding process to take place. In contrast, the other residues at the "reducing end" merely display transient interactions with the carbohydrate binding site.

9. The reported lactose- and 6'sialyllactose-functionalized glycodendrimers can be easily accommodated by Viscumin, providing enhanced binding. NMR experiments assisted by molecular modeling permitted the identification of the key sugar-lectin binding features, while the modeling protocol suggested a model for the possibility of multivalent interactions of these ligands with this lectin.

Publications

Ardá A., Blasco P., Varon Silva D., Schubert V., André S., Bruix M., Cañada FJ., Gabius HJ., Unverzagt C., Jimenez-Barbero J. Molecular recognition of complex-type biantennary N-glycans by protein receptors: a 3D view by using NMR. *J Am Chem Soc* **2013**, 135, 2667-2675. DOI 10.1021/ja3104928

Ardá A., Berbís MA., Blasco P., Canales A., Cañada FJ., Fernández-Alonso MC., Marcelo F., Jiménez-Barbero J. Recent Advances on the application of NMR methods to study the conformation and recognition properties of carbohydrates. *Carbohydr. Chem.*, **2012**, 38, 192-210.

Blasco P., Cañada FJ, Ardá A., Jiménez-Barbero J. and Mulard L. Molecular recognition of oligosaccharides related to the SF3a O-Antigen by monoclonal antibodies: a Three Dimensional view using NMR and MD simulations. **2014**, *Manuscript in Progress*